

Computational Study of Paxlovid in Ligand GA

Gordon Chalmers

Complex Carbohydrate Research Center

University of Georgia, Athens, GA, 30602

gordoncs@uga.edu, ORCID: 0000-0003-0254-352X

Submitted to Journal of Computer-Aided Molecular Design

Abstract:

Ligand GA is used to generate potentially realistic inhibitors of the 3CL protease of the SARS-Cov-2 virus by starting with the Nirmatrelvir molecule and changing it. After post-processing of the molecules using an extensive set of docking parameters, several of the molecules are selected and presented in detail that have good inhibition and metabolic activity, and binding increased to Mpro, while small change occurred in the binding to the liver enzyme CYP 3A4. This is a computational study of realistic potential inhibitors of the SARS-Cov-2 main protease found by using this software, with multiple results of which some are highlighted in particular. Of note is that these molecules may not require a CYP liver enzyme inhibitor such as Ritonavir in practical application. Keywords: drug design, SARS-Cov-2, machine learning, protein-ligand interactions, high performance computing, genetic algorithms, small molecule inhibitors, CADD

Section 1: Introduction

Ligand GA is a genetic algorithm (machine learning) tool recently developed [1] that produces small molecule ligands to proteins in a computational context by random changes of chemical structure (mutation) and recombination of existing structural fragments (crossover). The software, written in Matlab [2], creates optimistically high binding molecules based on the structure and chemical composition of the protein binding site. Notably, this is done without the use of external libraries of known small molecules or fragments, but rather with basic molecular manipulations of mutation and crossing over, guided iteratively by docking and scoring to a protein binding site and subject to ADME constraints. Ligand GA thus explores entirely a new and larger molecule search space. The fitness function in this machine learning approach is 0th order based on molecule structure and CCDC GOLD docking score but is changeable to include a variety of inputs. In this work, only ADME restrictions are used beyond the modeled binding interaction.

The genetic algorithm uses a SMILES textual representation of a molecule [3] as a chromosome. Corina Classic is used to convert the representations into mol2 or pdb files [4] and CCDC GOLD is used to perform the docking to protein cavities [5]. The population of chromosomes is computationally evolved from low scoring molecules to high scoring molecules. The landscape of the fitness function is very complicated due to atomic interactions between ligand and protein, and the algorithm will evolve molecules towards local or possibly global minima as the population of chromosomes iterate.

Considering that a small molecule can have dozens of atoms of different types and different modeled types of interactions (electrostatic, van der Waals, torsional energies, bond stretching, angular orbital overlaps, hydrogen bonds, solvent interactions, and others), this interaction and the maximization of binding energy is transcendental in nature. In addition, the protein amino acids in the docking cavity have side chains that are flexible, and the docking software CCDC GOLD [5] models this.

A potential realistic drug-like molecule cannot be found by docking score, i.e. binding interaction, alone. There are other considerations, such as tolerability in the body, synthesizability also at scale, and interactions of the molecule in the course of ingestion/injection towards the target protein(s), transport in the blood system, permeability in different media and through membranes, and observance of excessive binding to proteins in the biochemical pathways. ADME heuristics guided by Lipinski's Rule of 5 [6] and others [7-10], are a good measure of many of these, and almost all orally ingested small molecule drugs satisfy these guidelines. In addition to ADME requirements from Lipinski's Rule of 5, other constraints can be included from amino acid specific localization of ligand atoms, individual atomic binding scores, chemical fragments and substructure, pseudo-atoms, hydrogen bond constraints, etc... Only ADME constraints are used in conjunction with GOLD PLP docking scores in this work.

In [1] 2 molecules were presented, modifications of Aspirin bound to the COX-2 enzyme [11] and Simeprevir bound to the SARS-Cov-2 3CL Mpro [12] main protease, to demonstrate the use of the software Ligand GA. The analysis generated modifications of both molecules that have higher binding interaction and tighter fits to their respective protein cavities (binding sites), COX-2 and SARS-Cov-2 3CL Mpro. The modifications of Aspirin resulted in small molecules of >2x binding interactions and without any seeming structural oddities. Simeprevir to start out with is large and flexible, as were the Ligand GA modified molecules (ADME constraints were not used with that molecule's modifications).

In this work, a detailed study of Paxlovid [13,14] and its modifications is done. Paxlovid is an oral drug consisting of two molecules: Nirmatrelvir [13,14] and Ritonavir [15]. Nirmatrelvir is a small molecule of 35 atoms and satisfies all of the ADME constraints. Ligand GA uses this as the initial population in evolution to find better molecular bindings: better docking fit in the cavity, simpler structurally, better pharmacokinetic activity in the body in particular to the liver enzymes. One of the goals in this work is to produce small molecular modifications to reduce the metabolic rate of the liver eliminating it from the body; this in conjunction with significantly increasing the interaction with the Mpro site. Together this could eliminate the need for the added drug Ritonavir which has consequences of severely limiting Paxlovid's general use.

Background and Paxlovid

SARS-Cov-2 remains a continual threat to the population even after 2 years during which its molecular structure and its amino acid variations have been extensively documented [16]. Although vaccines have been successfully developed and used, there still remains the development of a viable therapeutic [17] that has no potential risk to the general individual. This is motivating in finding new small molecules, possibly unrelated to known molecules and without repurposing.

The search for orally ingested therapeutics caught attention recently after the successful but brief clinical trials of Merck's Molnupiravir [18,19] and Pfizer's Paxlovid. Paxlovid refers to Nirmatrevil in the presence of Ritonavir. Nirmatrevil, i.e. PF-07321332, is a small molecule that binds to a particular region of the SARS-Cov-2 main protease, blocking proteolytic cleavage of some of the SARS-Cov-2 polyproteins

[16], hence viral replication at an early stage of infection. The x-ray structure of Nirmatrelvir bound to the SARS-Cov-2 3CL protease, aka Mpro, is available at the Protein Data Bank, PDB ID 7SI9 [20]. The molecule is a descendant of an earlier created molecule, PF-00835231 [21], created during the SARS-Cov-1 outbreak in 2003¹. Nirmatrelvir's precursor is an injectable and was modified, turned into an orally ingested pill, which gained FDA EUA approval as Paxlovid in Nov 2021. However, the approval is for patients with high risk of hospitalization or death after infection. The inhibition of viral replication to avoid serious illness has a window of ~5 days post-infection, after that the virus has multiplied enough in the body that inhibition will not have an effect.

A serious drawback of Nirmatrelvir is that it metabolizes too quickly in the liver [13]. The cytochrome P450 class (aka CYP ###: gene family, subfamily, isoform/individual type) of enzymes is common, occurring in different species and contain a heme iron network that makes them function as oxygenases [22,23,24]. There are 57 encodings for CYPs in the human body, of which 6 are liver enzymes: CYP 1A2, 2C9, 2C19, 2D6, 2E1, 3A4, some of which are also found in the abdominal gut, and estimated to account for 80% of drug metabolism. These enzymes are the main source of metabolism and excretion of drugs in the human body. These, primarily CYP 3A4, eliminate Nirmatrelvir from the body and reduce the plasma concentration in the bloodstream.

Ritonavir is a well-known inhibitor of most of these enzymes and is commonly used to increase the concentration of other drugs in the body by non-specifically blocking the enzyme's normal metabolic function. It was initially developed as a protease inhibitor, in particular to HIV, but is now commonly used as an activator of other drugs by inhibiting the CYP enzymes, amongst other proteins. The molecule Ritonavir is large and flexible and sticks to almost any protein; it is not a precision designed small molecule, but the body can metabolize it. The problem with Ritonavir, and as a result with the Paxlovid therapeutic, is that by inhibiting these enzymes it also blocks the metabolism of a number of other molecules and drugs [25], including commonly used anti-depressants and cholesterol medicines. The use of Ritonavir could send a normal medication or dose of prescribed medication into a toxic level. There are a variety of side effects of Ritonavir [26]. As a result, Paxlovid is primarily intended for high risk groups in its FDA approval [27], which is about 2% of the population in the US.

Section 2: Docking calculations

Before the details of the docking analysis are presented for the molecules there are some comments. Binding of a ligand to a protein is often quantified by a single number using docking software from a top scoring pose. The physical process of a ligand-protein interaction can not be defined by a single number, although being a good quantifier of interaction. The potential interaction energy well that a small molecule occupies geometrically when bound has structure: depth, width, and high dimensionality coming from the many body atomic interactions of atoms from 2 molecules (electrostatic, van der Waals, torsional rotations, longer range hydrogen bonding, electronic quantum structure, and so on). There are generally different and complicated high-dimensional well shapes of energy minima and, as a result, different bound states reflected in docking software by different ligand conformations and perturbations (wiggles) of these. The stability of these bound states is dictated by characteristics of the local potential well, translating to evaluation of the disassociation constant k_D . This information is

¹ The molecule PF-00835231 was created in 2003, and is an early year in the area of modern computational chemistry in molecular simulation and modeling with computers. PF-07321332, Nirmatrelvir, is a modification.

required for accurate estimation of the on-/off- population of a ligand to a protein and its inhibitory effect.

In a multi-state system after sufficient time, the population of the ligands can be found by a usual thermodynamically Boltzmann weighted distribution. These primary states can be small in number, but with a spread in energy of almost degeneracy; thermal motion of a stable ligand conformation can result in near degenerate energies depending on the ligand and protein. The energy spread about these primary conformational modes can be visualized by physical wiggle room, e.g. spatial perturbations of a bound small molecule coordinates in an almost degenerate but continuous set of states trapped in a well while maintaining boundedness of the ligand. This spread depends on the ligand and protein.

GOLD is a non-deterministic genetic algorithm and running it multiple times samples the possible binding states, local and global minima of the binding interaction model fitness function. Large sets of docking jobs of Nirmatrelvir and all of the molecules presented in this work was done to sample the distribution of states quantified by docking score, i.e. binding interaction energy. The histogram of the docking scores of Nirmatrelvir/Mpro is shown in Figure 1. There are seemingly 2 dominant modes of ligand-protein interaction, a broad one at 46 and a narrow one at 68 GOLD PLP score. A finer bin resolution points to additional sub-structure about these scores, Figure 1(b) for example.

The distributions found by large sampling with random initial conditions from docking software can be useful in reducing the dimensionality of the dihedral angles of a ligand to protein in optimistic binding goals. A primary smeared peak in the energy density of states points to a reduction in the degrees of freedom in describing the ligand and its conformations near that point, such as at 68 and 46. In the case of the peak at 68, 5 of the 9 rotatable dihedral angles do not change much, and also perhaps 2 others. The 'lower half' of the molecule flip flops substantially in several different orientations within this set, and is due to only one degree of freedom angle. This is found by visualizing many of these molecules at ~68 docked score values.

A potential inhibitor of a protein, including that of the Mpro SARS-Cov-2 protease, has to have physical capability of entering into its docked pose. The small molecule should be able to rotate in dihedral angles and ring conformational states without obstruction to enter and bind. Docking software does not model this process before somewhat bound, but only the binding based on molecular models and modeling and physical data. Two widely separated peaks, without any scores in a histogram between them, are disjoint and can be interpreted – large variation geometrically of the molecule, although allowed and keeping its stereoisomer, are in different conformational states which are not smoothly connected. Two widely separated peaks with populated states in energy between, i.e. smooth (probabilistic) density of states, can be connected by small changes in the molecule and conformations. This is important in determining if a computationally determined molecule is reasonable. As an extreme example, an isolated cavity deep within a protein without any physical pathway outlets is unreachable, but could score very high or elsewhere in a distribution of scores via docking calculations.

These points are relevant to using a computational docking tool that measures conformational poses based on simplified energy calculations. As an example, Figure 2² shows two conformations of

² It is very important to emphasize the use of a molecular viewer such as Chimera [28], Pymol [29], VMD [30], or Nanome [31] to get a proper understanding of the 3-d protein-ligand complex. All of the .mol2 files of the figures

Nirmatrelvir/Mpro, both in the highest distribution peak of a large set docking calculations with random but chemically constrained torsion angles by modeled atomic interaction in GOLD. (The highest scoring pose at 71 is presented in next Section 3). These 2 superimposed example docking states have PLP scores of 68.5 and 68.2 and well representing the peak at 68 in Figure 1(a). The spatial difference of the two, i.e. conformations, is due to a single dihedral angle in the molecule, and the peak distribution at 68 is from the conformational change, quantified by the GOLD docking software and its initial conditions, guided by molecular interactions towards local minima. Most wiggles and small energy differences in the distribution come from small thermodynamic perturbations of geometric shape and orientation, not two different conformations near each other in binding energy.

Figure 1: Histogram of GOLD docking job scores of Nirmatrelvir to Mpro.

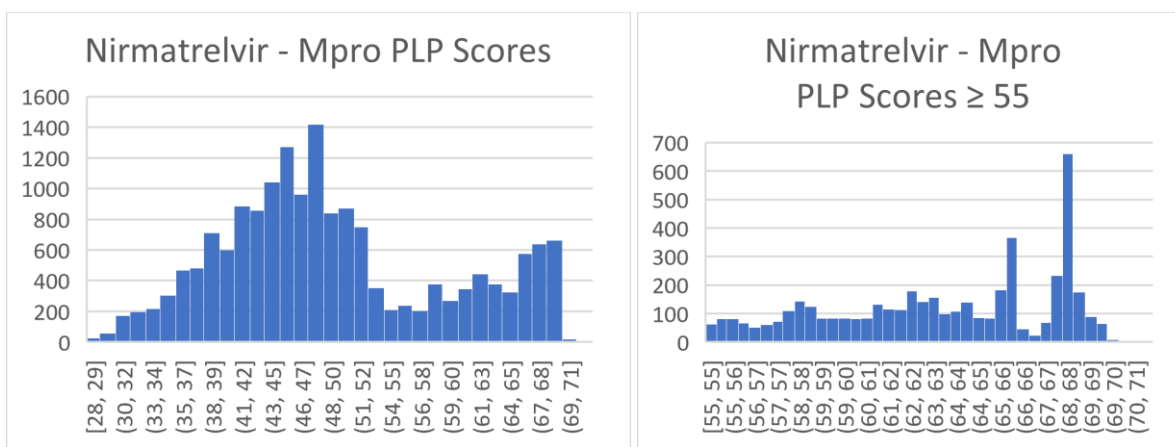
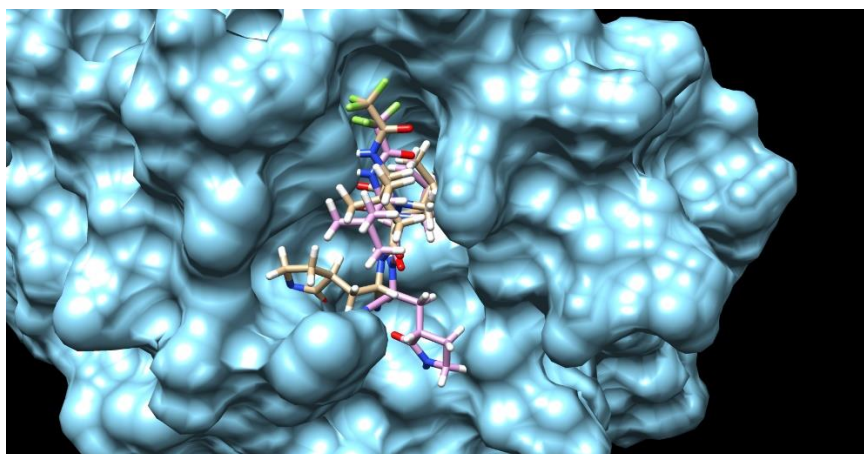


Figure 2: Two superimposed conformations of Nirmatrelvir/Mpro which differ by a rotation of one dihedral angle and scores 68.50 and 68.20. These scores are in the peak at 68 to 69 of Figure 1. Realistically, both states are active, being the same the stereoisomer of the same molecule.



are available in the Supplementary Information, and unused .mol2 files. There is a READ_ME_Figures.txt with the figure <-> file(s).

Energy distributions and density of states calculations via random sampling of binding states is necessary to find accessible binding conformations out of all possibilities in addition to filtering against geometrically realistic paths to the binding. These histograms are reported for the molecules examined in detail in later sections of this work.

The docking calculations are also performed on a primary drug metabolizing human liver enzyme P450, CYP 3A4. This enzyme is one of 6 CYP enzymes responsible for extracting waste-like molecules from the human body. CYP 3A4, while being considered the most prevalent, is one of these. The crystal structure of Ritonavir bound to CYP 3A4 [15] was used in the docking calculations, with the ligand stripped. The binding site of CYP 3A4 is a large vacuous chamber on one side of the heme iron atomic frame. CYP 3A4's binding site will stick highly to very general molecules. Because of its size geometrically, larger small molecules can easily 'fit' into this region without obstruction which is presumably why this enzyme is good at transporting waste molecules from the liver. A total ligand-protein binding energy will scale with the number of heavy atoms, so that a bound small molecule twice the size of another could result in a larger, even ~2x, larger binding energy. However, the target protein binding site of the drug is amino acid specific in function and increasing the size can miss the target and also create unwanted interactions with other proteins.

Section 3: Paxlovid

Paxlovid is a combination of 2 molecules: Nirmatrelvir and Ritonavir. Paxlovid in the literature is often referred to Nirmatrelvir in reference to the combination with the Ritonavir molecule.

Nirmatrelvir

The non-isomeric Nirmatrelvir molecule in SMILES notation is,

CC1(C2C1C(N(C2)C(=O)C(C(C)(C)C)NC(=O)C(F)(F)F)C(=O)NC(CC3CCNC3=O)C#N)C eqn. 1

This non-isomeric molecule has 32 stereoisomers, one of which is the Nirmatrelvir molecule,

CC1([C@@H]2[C@H]1[C@H](N(C2)C(=O)[C@H](C(C)(C)C)NC(=O)C(F)(F)F)C(=O)N[C@@H](C[C@@H]3CCNC3=O)C#N)C eqn. 2

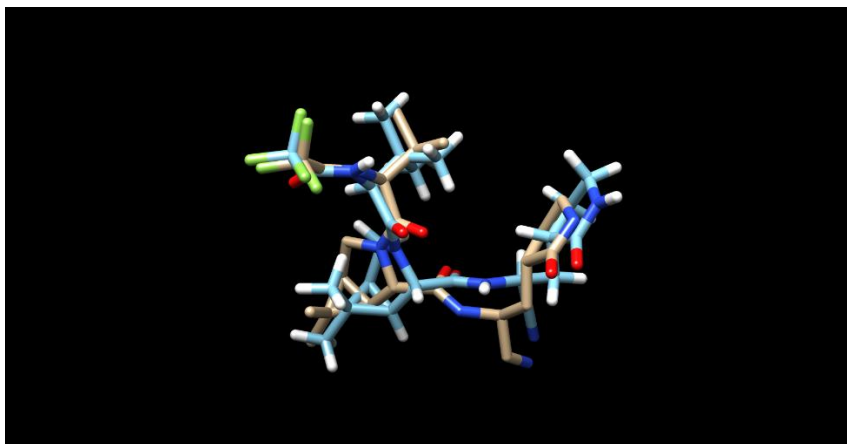
Nirmatrelvir has the molecular and ADME properties:

Mass (Da)	Heavy atoms	Rotat. bds	Hbond-don	Hbond-acc	Ow
499	35	9	3	9	

It is borderline in satisfying the well-known ADME requirements [6-10] given these characteristics.

The x-ray crystal structure PDB ID 7SI9 [20] of Nirmatrelvir bound the SARS-Cov-2 was used for docking and also in validating the docked pose. The computational docked pose is close to that of the complete x-ray structure, including the eclipsed trifluoromethyl group. The docked pose overlayed with the 7SI9 ligand is illustrated in Figure 3. The trifluoromethyl is in an eclipsed position which is unusual because typically this is higher in energy than non-eclipsed. Studies of these groups have been done with this eclipsed position and found to be more stable in classes of compounds [32]. This is the highest scoring docked Nirmatrelvir/Mpro pose generated by CCDC GOLD out of multiple runs. The RMSD of the 2 molecules in Figure 3 is 2.89 Angstroms as calculated in PyMol, comparable to the 2 Å x-ray resolution.

Figure 3: Nirmatrelvir from 7SI9 superimposed with the computationally docked molecule.



A Chimera view of Nirmatrelvir in its highest scoring docked pose, its Fischer projection in ChemDraw, and it bound to Mpro are in Figure 4. The distributions of docking runs to both proteins, Mpro and CYP 3A4, is in Figure 1. Several structural characteristics of Nirmatrelvir are: it has a fused amino-like double ring, a trifluoromethyl group which is eclipsed when bound to Mpro, and a nitrile extension.

The SARS-Cov-2 3CL protease (Mpro) binding site is an X-shaped cavity which can be described as: 3 ends are caves and the fourth is an exiting valley. This is the region of the Mpro relevant to viral replication which is targeted and partly blocked by Nirmatrelvir. This molecule, when bound, primarily fills in 2 caves.

Note: In this image of Nirmatrelvir bound to Mpro, the solvent accessible protein surface (SAS) is displayed. This has the advantage of clearly showing the binding cavity. However, in this image and in general there could be a hydrogen appearing to penetrate the surface. This is the case with Nirmatrelvir and Mpro in the nitrogen atoms 19, 36 farthest, nearest nitrogens from the trifluoromethyl group. These both are hydrogen bond donors (info on next page) with Mpro atoms 1315, 1112 in hydrophilic GLU 166, hydrophobic PHE 140. There is no structural clash or overlap, and this is common in ligand-protein Chimera images. Figure 5 shows the ligand binding without the surface.

The docking information of highest scoring Nirmatrelvir to Mpro and CYP 3A4 is

	Score	S(PLP)	S(hbond)	S(cho)	S(metal)	DE(clash)	DE(tors)	Intcor
Mpro	70.68	-64.58	3.00	0	0	0	1.60	.28
3A4	84.44	-82.76	1.00	0	0	0	1.00	0.68

This particular state of Nirmatrelvir is an outlier in the histogram of docking runs. First, it is the highest scoring conformer. Second, there are 11 out of these that scored rounded to either 70 or 71. That is a PLP score of 2 difference, 2 steps of discreet jumps at ~1 to 69, which is unusual and not typical of wiggle in conformations. This gap in states, although not blatant of a total 1/3 kCal, could be due to the molecular interaction modeling. There doesn't seem to be anything conformationally that would make this conformer to stick out from the bump at 67-69 in Figure 1.

The geometric pose agrees with the x-ray crystal structures made with Mpro-Nirmatrelvir [17]. X-ray samples of protein-ligand complexes are unique in energy minimization however, and the very much larger density of states at 67-69 with a spread of ± 3 , Figure 1(a,b) is biologically realistic, with a typical score of 68.

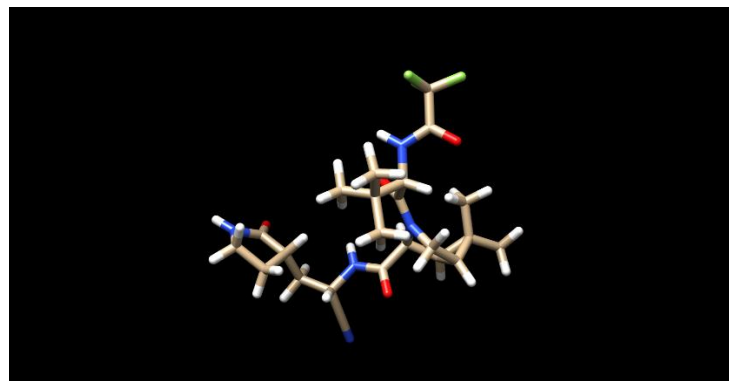
There are 4 ligand-protein hydrogen bonds,

Donor molecule	Donor type	Donor heavy atom	Donor atom	Acceptor molecule	Acceptor atom	Score	Amino acid
L1	H	19	56	P1	1315	1.00	GLU 166
L1	H	36	64	P1	1112	.61	PHE 140
P1	H	1136	2470	L1	40	.60	GLY 143
P1	H	1312	3711	L1	12	.91	GLU 166

The difference in score from Nirmatrelvir-Mpro to Nirmatrelvir-CYP 3A4 is $84.4 - 70.7 = 13.8$. This is very large, approximately of 2.1 kCal in binding energy. GOLD PLP scores are dimensionless numbers representing the interaction of docked ligand to the protein; the rough correspondence between PLP score differences and energy differences is found by known calculations in terms of both and is roughly 6.5 PLP for 1 kCal of various molecules. The binding energy is important and biochemical terms of ligand in the on-/off- states of protein-ligand is important. In terms of k_D or on-off ligand binding fraction, 1 kCal is commonly referred to be a factor of 10, although this is not a linear or rigorous relation.

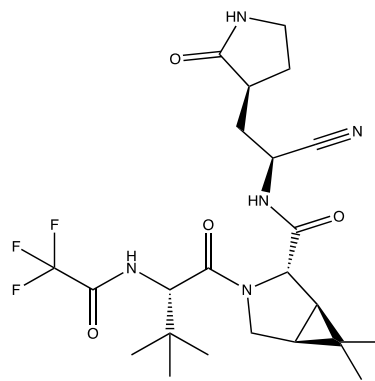
Figure 4: (a) Nirmatrelvir in its highest docked pose to SARS-Cov-2 Mpro. (b) Diagram of Nirmatrelvir. (c) Nirmatrelvir bound to Mpro. (d) Nirmatrelvir bound to CYP 3A4.

(a)



(c)

(b)



(d)

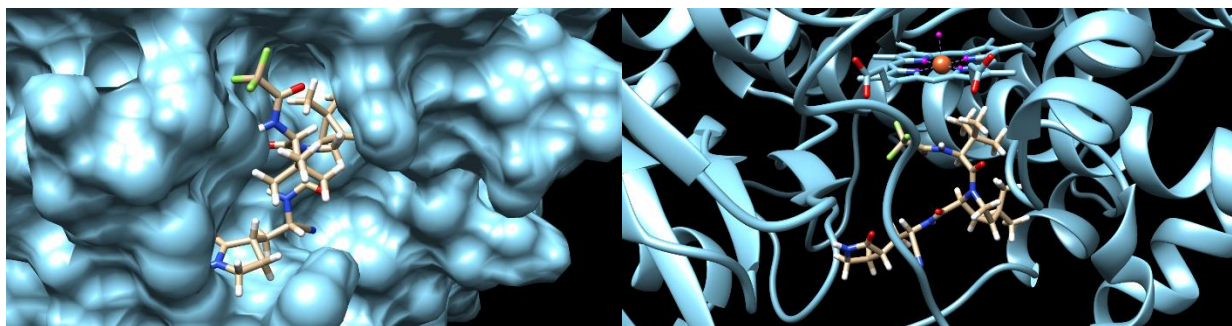
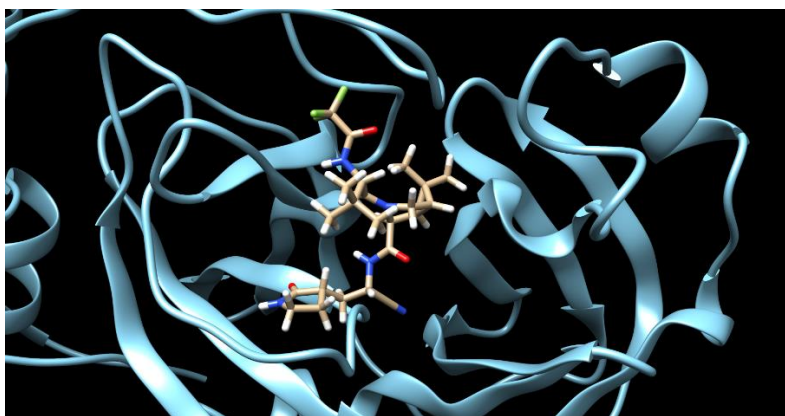


Figure 5: Nirmatrelvir bound to Mpro without the SAS displayed.



Studies of the metabolism of Nirmatrelvir revealed that it doesn't last for long in the body. The liver enzymes are actively working fast with it. As a result, plasma concentrations of it decrease too rapidly. 2 possible solutions to maintaining sufficient presence to have an inhibitory effect on viral replication via the target Mpro are: 1) use higher and more frequent doses, 2) add an inhibitor to the garbage collection process by blocking the action of the CYP enzymes, in particular 3A4. Both have drawbacks: 1) unwanted interactions with other proteins, 2) blocking the liver enzymes from normal function could result in increased levels of other medications to the point of toxicity. Dose changes for either purpose obviously means more interactions with a variety of proteins, not only Mpro.

Ritonavir

Ritonavir (Norvir) is a well-known inhibitor of CYP 3A4, and most of the CYP liver enzymes. Initially developed for treating HIV, it became well known as a CYP inhibitor and is currently used in conjunction with other drugs to maintain or raise their concentrations in the body. It is a very large small molecule that breaks most of Lipinski's Rule of 5 ADME conditions. It has mass of 720 Da and 18 rotatable bonds. It is so flexible and large that it potentially sticks to almost anything; it is not a precision small molecule inhibitor of a protease. Ritonavir is also a protease class specific inhibitor and was considered in conjunction with Liponavir as an inhibitor of Mpro due to latter's structure. Many molecules of this size and flexibility have been both computationally and in the lab examined as possible viable inhibitors of SARS-Cov-2 Mpro [33,34].

The canonical representation of Ritonavir is,

CC(C)C1=NC(=CS1)CN(C)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CC2=CC=CC=C2)C[C@@H]([C@H])(CC3=C
C=CC=C3)NC(=O)OCC4=CN=CS4)O eqn. 3

Molecular properties of Ritonavir are:

Mass (Da)	Heavy atoms	Rotat. Bds	Hbond-don	Hbond-acc	ow
720	50	18	4	9	6

Ritonavir was docked both to Mpro and to CYP 3A4. Docked to Mpro and CYP 3A4

	Score	S(PLP)	S(hbond)	S(cho)	S(metal)	DE(clash)	DE(tors)	Intcor
Mpro	100.37	-99.24	2.90	0	0	3.85	2.61	1.49
3A4	142.18	-133.97	2.00	0	.99	0	3.24	2.77

The molecule, diagram of it, its highest docking pose to Mpro, and its highest scoring pose to CYP 3A4 are shown in Figure 6. The distributions of docking runs is given in Figure 7. The distribution is wide and very normal; this is due to the largeness and also the flexibility of 18 rotatable bonds giving it the ability to stick in almost any conformation. The peaks,widths pertaining to Mpro and CYP 3A4 are centered at 57,10 and 102,11.

The docking of Ritonavir to CYP 3A4 results in an extremely high PLP docking score, 142. This confirms its use as an effective inhibitor. It essentially fills the entire cavity adjacent to the heme iron frame, see Figure 6. Note that it has 50 heavy atoms and its size is primarily responsible for the high PLP docking (interaction) score. Per atom the score is 2.56, which is high, but the filling of the cavity and the number of atoms is the primary reason for its large interaction with CYP 3A4. By contrast, Nirmatrelvir has 35 atoms, a score of 84.11, and a similar score per atom of 2.41 to CYP 3A4.

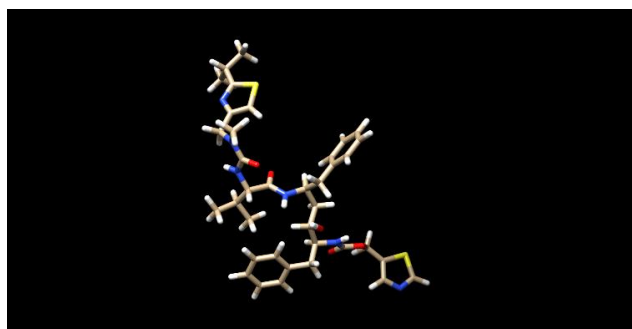
The SARS-Cov-2 Mpro is structurally similar to Ritonavir, and Liponavir-Ritonavir, has been examined in inhibiting Mpro, the latter of which is in the class of proteases that it binds to. The docking score is 100.37, which appears reasonable when compared to Nirmatrelvir's score of 68 or 70.7. However, as mentioned, Ritonavir is large and flexible and can generally interact and fit somewhat to general proteases and proteins. Upon inspection of the docked Ritonavir to Mpro, it extends into one X-cave and the surface tunnel. It partly enters another X-cave and covers over, not in, the 3rd cave. It possibly enters the target region necessary to inhibit biochemical viral replication. This is a good example of why localization of small molecule to the protein is more important than a high docking score in realistic molecular design, however both being important. Ritonavir was found in conjunction with Liponavir, and also in other combinations, not to be an effective drug in blocking the SARS-Cov-2 [35].

Ritonavir's use in blocking the metabolism of other drugs is well known. However, its large effect of 'stuffing the waste basket until full' of CYP 3A4 (and other CYP liver enzymes) so that that a molecule can not be cleared by the liver has the disadvantage of blocking a variety of normal liver functions and of other drugs/small molecules. This a reason that the drug Paxlovid is FDA approved only for high risk groups, i.e., those individuals with predisposition to severe Covid-19, who can accept the risk of

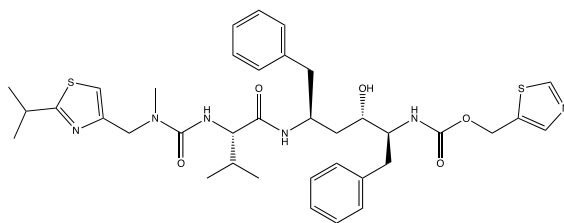
Ritonavir. This disadvantage could be eliminated or alleviated by finding other potential small molecule candidates that don't require Ritonavir or other drugs to inhibit ordinary bodily function.

Figure 6: (a) Ritonavir in its highest docked pose to CYP 3A4. (b) Diagram of Ritonavir. (c) Ritonavir bound to CYP 3A4. (d) Ritonavir bound to Mpro.

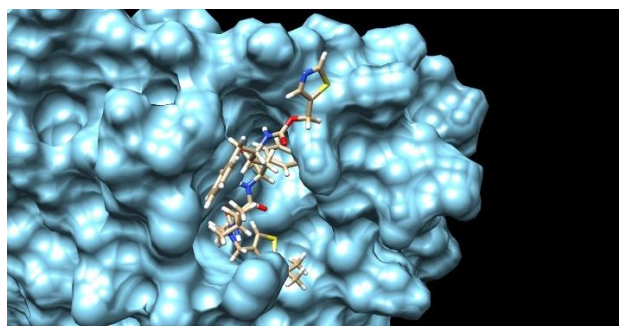
(a)



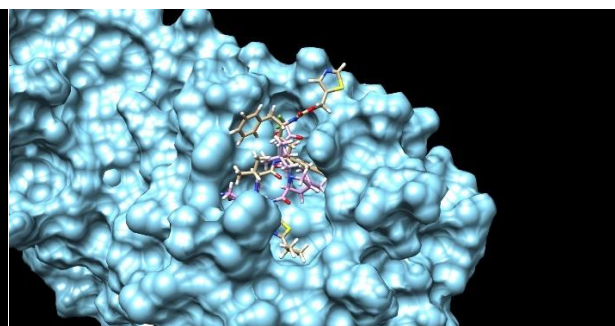
(b)



(c)



(d)



(e)

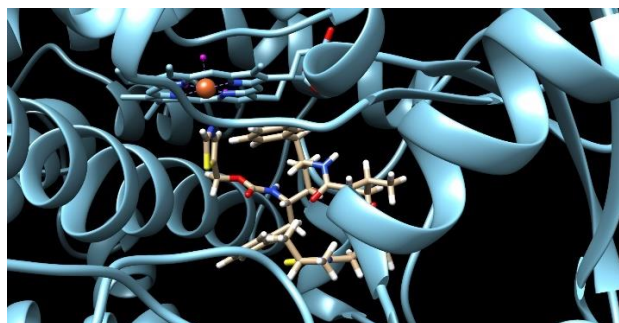
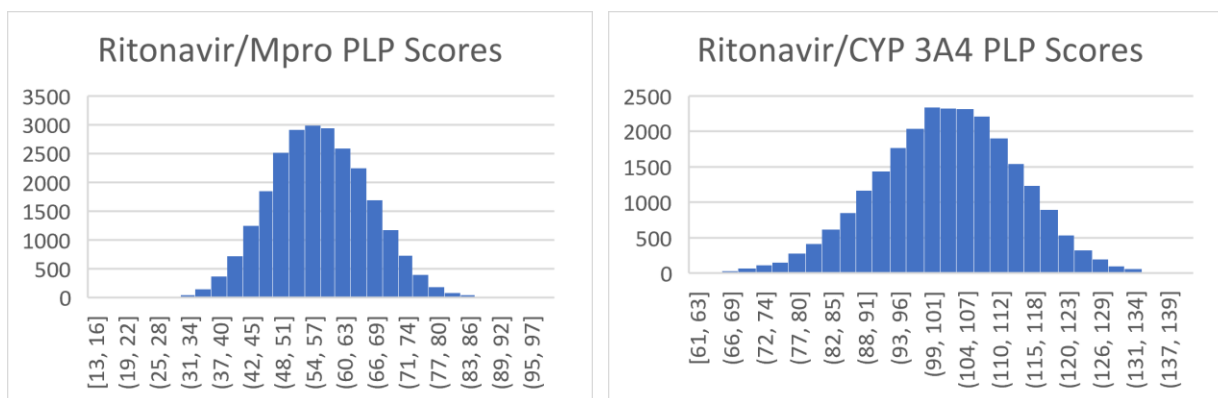


Figure 7: Distribution of (a) Ritonavir/Mpro scores. (b) Ritonavir/CYP 3A4 scores.

(a)

(b)



Section 3: Computations

A lengthy run of Ligand GA was used to produce new inhibitors of the relevant SARS-Cov-2 main protease binding site. The genetic algorithm was initialized with 30 identical non-isomeric Nirmatrelvir molecules

A single node of 64 cores with a total of 128gb of RAM was used at the GACRC for the Ligand GA calculations [36]. The total computing time of Ligand GA was ~150 hours with 33 iterations of the Ligand GA genetic algorithm. In order to understand the complexity consider 30 non-isomeric molecules each with 128 stereoisomers. In the Ligand GA fitness function, the non-isomeric molecule is expanded onto all stereoisomers, each of which have to be docked. The fitness function included ADME penalties. The average over the course of 30 iterations is ~200 stereoisomers, 15 runs each, and 30 iterations of 30 of these resulting in total docking jobs of ~3M using default parameters apart from 15 runs and the default of auto-scaling of 1.0. The total time required for a docking calculation varies depending on the complexity of the molecule, binding site, and parameters used in CCDC GOLD.

In the post-processing of the 15922 molecules there was an average of 210 stereoisomers per non-isomeric molecule and an average 6285 docking calculations. Each post-processed docking job took an average of ~400 and ~600 seconds (in Mpro and CYP 3A4) because 30 repeated calculations were done. This used about 40 hours on a single 64 core node. The final step of post-processing of 4 small molecules and 2 proteins used Cluster GOLD with multiple nodes at ~30 hours.

GOLD Parameters

There are 2 sets of parameters, from Ligand GA and from CCDC GOLD. The GOLD parameters during the iterations were chosen with mostly auto default: 15 identical docking jobs were run in each GOLD molecular docking, an autoscale=1.1, early termination was turned on, and a cavity.atoms file was used with a radius of 20 Angstroms centered about. After the ranked set of non-isomeric files was obtained after 4 or 5 iterations using Ligand_GA_Output_Analysis, Ligand GA was restarted with the 30 top scoring non-isomeric molecules. This continued until 33 iterations were completed. Then the net set of the result from the total collection was ranked and ordered using Ligand_GA_Output_Analysis. A truncated list of 76 molecules was selected based on any GOLD PLP score greater than 79; this cutoff was chosen because the highest score of any of Nirmatrelvir's stereoisomers is 78. Then this list of 76 non-isomeric molecules was fully post-processed with an extensive set of GOLD parameters: 30

repeated docking calculations per ligand, pop_size=75, n_islands=10, max_ops=300000, early_termination=0 (off). The set 76 non-isomeric molecules is given in the Supplementary Information. The resulting number of molecules is 15922, coming from 76 non-isomeric molecules (average 260 stereoisomers). The total number of post-processing GOLD runs is ~600,000.

The GOLD docking parameters during the Ligand GA run in using mostly defaults are for faster iterations at cost of less accurate evolution. Due to GOLD being a genetic algorithm also, the results can slightly vary because a genetic algorithm is not necessarily deterministic. Scores can change if not enough search is made in GOLD but in repetition the docking random sampling of initial dihedral angles will stabilize and show the conformational modes of the ligand to the protein.

Ligand GA parameters

The point of stopping after 4 or 5 iterations and then restarting with a ranked list is to improve the rate of population improvement towards a local minima. This is a typical maneuver in using genetic algorithms if the fitness function is very complicated and the time of computation required very large. A population of 30 chromosomes is sufficient to obtain results in 33 iterations. A typical iteration requires about 4 hours with 64 cores. Genetic algorithms are known to converge rapidly in less than 15 iterations, very much depending on the complexity of the fitness function and size of the search space, and much more iterations are necessary to improve accuracy at the minima, i.e. flattening off behavior of genetic algorithm best scores. It should be noted that due to the non-zero elite_fraction in the GA, a portion of the non-isomeric molecules is resampled and docked by GOLD, and this improves the accuracy of the calculation while others are kept after the 4 or 5 iterations.

The molecules were evolved in the genetic algorithm Ligand GA software by using the SARS-Cov-2 Mpro binding site. In addition, the molecules were post-processed on a primary liver human enzyme CYP 3A4. The same extensive GOLD parameters were used on both proteins for all molecules, and a 20 Angstrom sphere cavity file was used with the sphere centered on the Mpro GLU amide and the CYP 3A4 heme iron.

The set of Ligand GA parameters used was the default set described in [1] with 2 changes,

- uniform_percent_atom=[1,1,1,.5,.5,.5]
- max_tries=200
- unnorm_mut_probability=[1,1,1,1,1,1,1,1,.4,.4,.4]; (and with .4 -> .2 in final 3 iterations)

Matlab GA parameters used by Ligand GA are,

- crossover=.9, mutation=.9
- elite_fraction=.15

with a population of 30 individuals. Convergence criteria was set to run until stopped, which was done every 4 or 5 iterations, about 3 hours each, and for repopulating the next run of 30. The input files for Ligand GA with all of the parameters are in the Supplement along with the input files to run GOLD.

These parameters gave a weight for the atoms used C,N,O,P,S,F of 2x for C,N,O than P,S,F. The weight of single-double, single-triple, and vice-versa were .4 any other mutation probability. The crossover and

mutation rates are .9 and elite_fraction=.15 due primarily due to the complicated docking function, the interatomic interaction energy modeled by CCDC GOLD.

There is much information in a gold docking output mol2 file. The GOLD output mol2 files not only carry the coordinates and conformation of the poses but also detailed ligand and amino acid information, bond information, amino acids in the cavity, and other information. *Docking score values are from modeled interactions of true inter-atomic interactions and are not true to 2 significant digits. The numbers are kept to this precision to reflect GOLD output and all distributions used rounded integer numbers of score values S(PLP)). The files in Supplementary Information have these values to 2 significant digits.* The overall docking score summary of each molecule in this work is quoted in terms of:

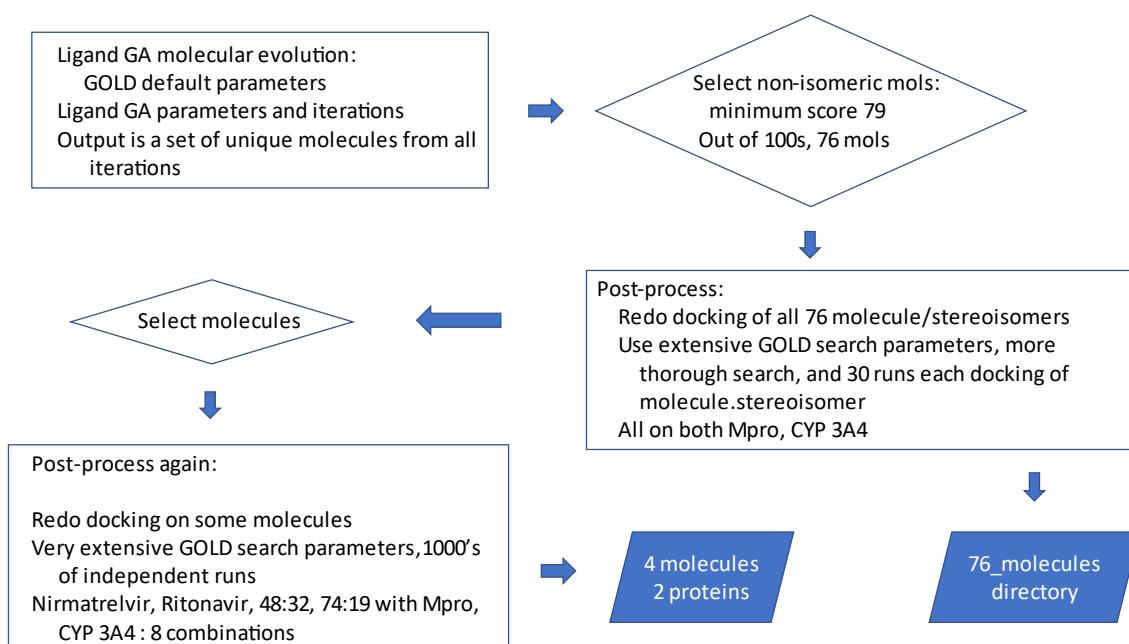
S(PLP)	S(hbond)	S(cho)	S(metal)	DE(clash)	DE(tors)	Intcor
--------	----------	--------	----------	-----------	----------	--------

Individual GOLD output ligand PLP atomic scores can also be requested by changing GOLD parameters, and is useful in editing molecules. In mention, the Ligand GA fitness function was changed to average docking score per heavy atom, and several multiple iteration runs were done with this but not presented.

The Ligand GA fitness function uses an S(PLP) - ADME_Penalty for each of the molecules and stereoisomers. ADME_penalty enforces the Lipinski Rule of 5 conditions in the soft sense [1]. There are additional constraints that have been written for a completely different set of molecules and goal, including enforcing particular hydrogen bonds, but these are not used in this analysis.

The calculations in this work follows the procedure in the flow diagram in Figure 8. The GOLD parameters are different in the 3 stages: default 15 runs, extended search parameters 30 runs, extended search 200 runs. The GOLD files are in the calculations directory of the Supplement.

Figure 8: Logical flow of using Ligand GA and post-processing.



Step 1 uses Ligand GA. Then by using a cutoff of 79 for each of the non-isomeric molecules, 76 molecules are used in step 2. In step 2, these 76 molecules are processed with a more extensive set of parameters,

- 30 repeated docking calculations per ligand
- pop_size=75
- n_islands=10
- max_ops=300000
- early_termination=0 (off)

Then from these results, an even more exhaustive set of calculations is performed for molecules 48:32 and 74:19, and Nirmatrelvir, Ritonavir, consisting of 100 identical molecules and 200 GOLD docking runs. All of the processing after Step 1 uses both Mpro and CYP 3A4.

Any of the molecules at Steps 1-3 can be edited for improvement in Ligand GA and also for realistic pharmacokinetics. This wasn't done the molecules used in this work. The different output directories and files can be used to reproduce any of the calculations.

Section 4: Analysis

There are several criteria in searching the set of 15922 for good choices:

1. Increased docking score, i.e. binding interaction, of molecule/Mpro compared to Nirmatrelvir/Mpro
2. Decreased docking score of molecule/CYP 3A4 compared to Nirmatrelvir/CYP 3A4
3. Higher score to Mpro to CYP 3A4 is preferred.

- Satisfactory visual geometric docking fit to the site and individual atom interactions between protein and ligand.
- Synthesis, maybe at scale, is important and is analyzed by inspecting individual molecules.³

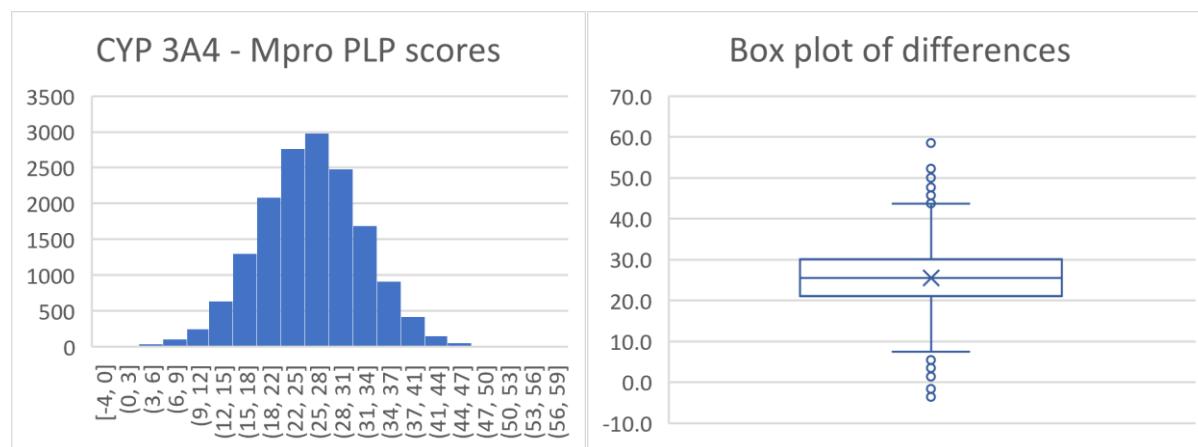
The iterations of the molecules in Ligand GA follows the highest scoring stereoisomer of the non-isomeric molecule. Molecules are represented in textual non-isomeric SMILES. Upon expansion of one of these many stereoisomers will be generated, and the highest scoring one is what guides the selection from one iteration to the next of the non-isomeric population.

First the results from the 15922 molecules are given, of which Nirmatrelvir is not included because its score is less than 79. There are more than 15922 in the complete output, but a cutoff slightly higher than the largest score of the stereoisomeric completion of Nirmatrelvir was used to limit the analysis and post-processing of the total set. Figure 9 gives the distributions of these molecules, in score differences between molecule/Mpro to molecule/CYP 3A4, the scores of molecule/Mpro, and the scores of molecule/CYP 3A4.

Figure 9: (a) Histogram of CYP 3A4 – Mpro PLP score differences. (b) Whisker plot. (c) Histogram of molecules docked to Mpro. (d) Histogram of molecules docked to CYP 3A4.

(a) mean, median, std = (25.5, 25.5, 6.9)

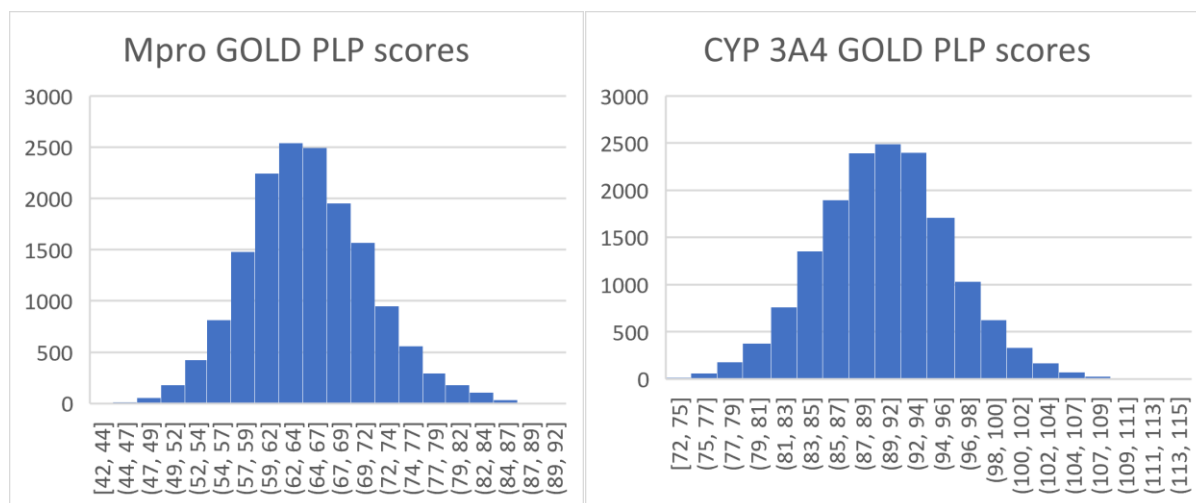
(b) mean, median, std = (25.5, 25.5, 6.9)



(c) mean, median, std = (64.8, 64.4, 6.4)

(d) mean, median, std = (90.3, 90.3, 5.4)

³ The MIT Machine Learning for Pharmaceutical Discovery and Synthesis Consortium (MLPDS) <https://mlpds.mit.edu> is dedicated to analyzing structures for synthesis planning, likelihood, and other related information. There is a site <https://askcos.mit.edu> that can be used to determine possible synthesis pathways of molecules, including from SMILES format [37].



The mean, median, and standard deviation of the molecules are:

	Mpro scores	CYP-3A4 scores	CYP 3A4 – Mpro scores
Mean	64.8	90.3	25.5
Median	64.4	90.3	25.5
Standard Deviation	6.4	5.4	6.9

The PLP score of 70.7 of Nirmatrelvir is on the border of the 1st quantile (71.2) of the total set of 15922 stereoisomers. It's score 84.4 to CYP 3A4 is at the border of the 2nd quantile of scores (84.9).

The Mpro/CYP 3A4 score Pearson coefficient of the 15922 molecule scores docked is .33, a low correlation between binding strength of the molecules to 2 proteins. The 2 binding sites are from 2 proteins and are very different, and a high binding to Mpro with low binding to CYP 3A4 is a goal of the analysis. CYP 3A4 has a large and bland cavity in which molecules bind into; it isn't surprising that there is no correlation.

The average difference of the Mpro and CYP 3A4 scores is 25.5, approximately 3.9 kCal if 1 kCal per 6.5 is used. This high average can be problematic in the design of realistic drugs as very quick metabolism will result if it is much higher than the binding to the target protein. The difference of Nirmatrelvir is 13.8, which could be improved upon.

Selection of molecules:

Selecting particular molecules depends on the criteria. All of the molecules satisfy Lipinski's ADME restrictions as the molecules were evolved with an ADME penalty in the Ligand GA fitness function. The conditions 1,2,3,4 are used to present 2 sets of molecules. The docking of molecule to CYP 3A4 is not used in Ligand GA but analyzed after the molecules were generated. (Both proteins are used simultaneously in the to be released version Ligand MultiObj GA).

Out of the 15922 isomeric molecules generated from the 76 non-isomeric molecules, 5 had docking score difference between Mpro and CYP 3A4 that is positive (criteria 3). All of these had higher

ligand/Mpro scores than Nirmatrelvir/Mpro, and 4 had lower scores to Ritonavir/CYP 3A4 (criteria 1,2 together with 3 is marked by an x in Table 1). The molecules are listed in Table 2 and their molecular representations are in Table 2. The mol2 files and GOLD docking output are given in the Supplementary Information. Two of these molecules, 48:32 and 74:19 are discussed in detail, and the molecular diagrams of the other 3 molecules, stereoisomer specific, are shown in Figure 10.

Table 1: 7 selected molecules based on difference scores and criteria 1,2,4 in first post-processing, (Figure 8).

mol_idx	steiso	diff		Mpro PLP	CYP 3A4 PLP		# heavy atoms
Nirma		15.4		67.93	83.28		35
37	268	-0.11		78.59	78.48	x	37
38	64	-0.53		80.39	80.92	x	28
48	32	-3.57		82.26	78.69	x	28
68	51	-0.24		83.65	83.41		36
74	19	-2.01		90.25	88.68	x	37

Table 2: 5 molecules from Table 1 in SMILES representation.

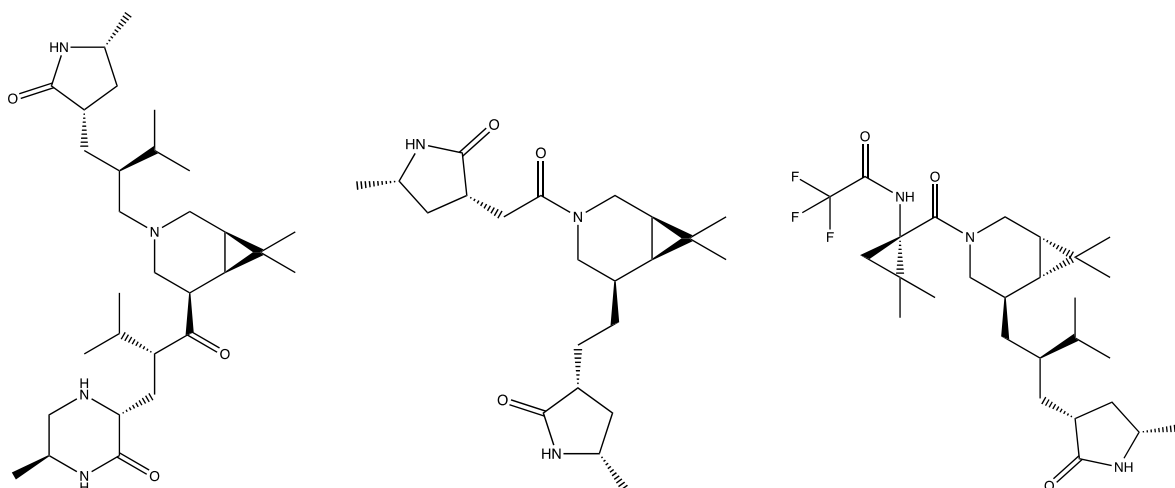
		Molecule
37	268	<chem>CC1(C)[C@H]2[C@@H]1[C@@H](C([C@H](C[C@H]3NC[C@@H](NC3=O)C)C(C)C)=O)CN(C[C@@H](C[C@H]4C[C@H](NC4=O)C)C(C)C)C2</chem>
38	64	<chem>CC1(C)[C@H]2[C@@H]1[C@@H](CC[C@@H]3C[C@@H](NC3=O)C)CN(C(C[C@@H]4C[C@@H](NC4=O)C)=O)C2</chem>
48	32	<chem>CC1(C)[C@H]2[C@@H]1[C@@H](CC[C@@H]3C[C@H](NC3=O)C)CN(C(C[C@@H]4C[C@@H](NC4=O)C)=O)C2</chem>
68	51	<chem>CC1(C)[C@@H]2[C@H]1[C@@H](C[C@H](C[C@@H]3C[C@@H](NC3=O)C)C(C)C)CN(C([C@@]4(NC(C(F)(F)F)=O)C(C)(C4)C)=O)C2</chem>
74	19	<chem>CC1(C)[C@@H]2[C@H]1[C@@H](C([C@H](C[C@@H]3C[C@H](NC3=O)C)C(C)C)=O)CN(C([C@@]4(NC(C(F)(F)F)=O)C(C)(C4)C)=O)C2</chem>

Figure 10: Molecules from Table 2. The remaining 2, 48:32 and 74:19, are described in detail.

a) Molecule 37:268

(b) Molecule 38:64

(c) Molecule 68:51



There are many other possibly viable molecules. Docking score, or measuring binding interaction, depends on the molecular modeling and is only -one rough- quantifier of potential binders. Although CCDC GOLD does produce x-ray validated docking of ligands, the fit of the molecule in the cavity and at the relevant region at amino acids is important. Ritonavir has a high docking score to Mpro, was considered as a potential inhibitor once, then experimentally found to have no effect; it does stick to Mpro but it is large and very flexible – it sticks away from the region of Mpro that is responsible for cleaving polyproteins in the viral replication process. The size of the molecule, i.e. mass or number of heavy atoms, is also important in considering possible inhibitors; increasing the docked score per atom is independent of the number of atoms; a total binding interaction does increase generically with the this number.

More molecules are listed in Table 3 which follow the conditions 1,2, and 4. Out of the complete set of 15922 there were 76 which satisfied conditions 1,2,4. Also, docking calculations is not an exact calculation, but being based on a genetic algorithm with dihedral angles is non-deterministic. A cutoff factor of 1.02 was used to multiply the score of the Ligand GA molecules in comparing with Nirmatrelvir. This conservatively raises the docking score of the molecule to both Mpro and CYP 3A4, and decreases the number of screened molecules than without (122 with 1.0 cutoff factor).

Table 3: Selected molecules based on criteria 1,2, and 4.

mol_idx	steiso	diff		Mpro PLP	CYP_3A4 PLP	# heavy atoms
Nirma		15.4		67.93	83.28	35
38	40	1.5		75.4	76.9	28
67	86	2.33		78.87	81.2	37
65	78	7.11		83.62	90.73	34

The set of 76 molecules all have a stereoisomer (15922 total from the 76) that has a higher binding score than Nirmatrelvir while satisfying ADME restrictions. There are many molecules out of 15922 with higher scores, as seen in the distribution Figure 9(c).

The 76 non-isomeric with scores above 79 are a direct output of Ligand GA. It is reasonable and important to manually edit molecules if necessary. For example, Molecule 74:19 and 68:51 have an external 3-member C ring that appears peculiar and raises the question ‘what is that for’. It does have an effect in the geometry of the stereoisomer and interaction, but it could be changed and then re-analyzed. Manual modifications of functional groups or atoms is expected in realistic inhibitor design. Simeprevir, a very effective treatment of Hepatitis C, has 2 3-member C rings and one of them is external.

Improvement of Nirmatrelvir so that interaction is increased with Mpro and decreased with CYP 3A4. In this work only Mpro was used in molecular evolution with a GA. The situation improves after using a multi-objective genetic algorithm. In Ligand MultiObj GA the small molecules are iteratively improved simultaneously using 2 fitness functions in the evolution, maximizing interaction with Mpro and minimizing with CYP 3A4.

Analysis of Molecule 48:32

A detailed analysis of one of these molecules, 48:32, is given.

The interaction of a ligand to a protein and biochemical behavior is not found by only a score, but rather a complete examination of the ligand, including visualization of the protein-ligand complex. The discussion begins with the highest docking scores of Nirmatrelvir and Molecule 48:32 to Mpro and CYP 3A4,

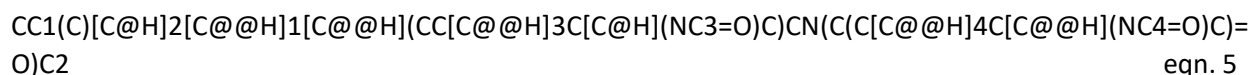
	Nirmatrelvir	Molecule 48:32
Mpro	70.68	82.61
CYP 3A4	84.44	86.55
Difference	13.76	3.94

This molecule fits in the set of 15922 as an outlier in the binding interaction score differences of Mpro and CYP 3A4, and in the highest quartile of scores to both proteins. Originally this molecule was selected on from duplicated set of GOLD runs with exhaustive GOLD search parameters on Mpro and CYP 3A4, which led to highest scores of 82.26 and 78.69, a difference of -3.57. Given the large number of multiple runs of docking job with different random initial conditions, from the 2nd post-processing step of 4 molecules (Figure 8), the highest Mpro scores changed slightly to 82.61 and 86.55 from that in Table 1.

Molecule 48 stereoisomer 32 is in terms of non-isomeric SMILES,



and isomerically,



It's molecular properties relevant to ADME are:

Mass (Da)	Heavy atoms	Rotat. Bds	Hbond-don	Hbond-acc	ow
354	28	6	2	6	?

Molecule 48:32 is 70% the mass of Nirmatrelvir. It has 80% the number of heavy atoms, and it has 60% the number of rotatable bonds than Nirmatrelvir. This molecule is much smaller and less flexible. It has 4 rings not 3, which makes it more compact in size. These parameters are seemingly better as far as tolerance and permissivity of it in the biological process of ingestion to the target protein.

The distribution of binding modes is in Figure 11 from 1000's of docking runs with exhaustive search parameters. There appear to be 4 primary Molecule 48:32/Mpro docking modes at 52, 60, 73, and 82. The score distributions (representing binding energy) and their widths in the histogram should be taken into account in comparing computationally the ligand binding of Nirmatrelvir versus molecule 48:32. All of the modes of molecule 48:32 are higher in binding score (energy) comparatively, with the 83 one from 48:32/Mpro much larger. The number of rotatable bonds is 6 and comparing to Nirmatrelvir's 9 is a 33% reduction. There are fewer rotatable degrees of freedom; the fewer rotations could be a source of the differences such as 4 major center points of its PLP score distribution in Figure 11, as compared to 2 of Nirmatrelvir in Figure 1.

The distribution of scores of Molecule 48:32/CYP 3A4 is useful in gauging how it binds both in value and conformationally to CYP 3A4. This distribution is in Figure 12. The highest score of 86.55 is an isolated point and the number of ligand conformationally bound states is far greater near the value of 79. Due to the seeming discontinuous nature of the distribution, following from a conformational change in the set it isn't clear if the actual binding is in fact greater to Mpro than to CYP 3A4. However, the drastic drop in number of available states after 79 shows a difficulty in rotational degrees of freedom at the interaction level.

Docking scores aren't exact and this molecule is a good choice to examine based on the scores and distributions, both to Mpro and CYP 3A4. The docking results (output files) of all the 48:32 docking jobs are available in the Supplementary Information and can be used to examine the conformational aspects and distribution of scores.

Figure 11: Histogram of GOLD docking job scores of Molecule 48:32 to Mpro.

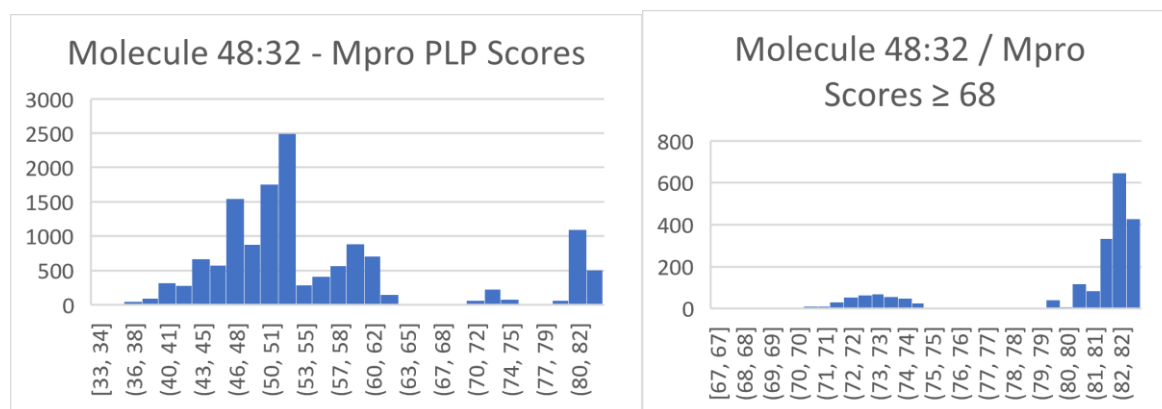
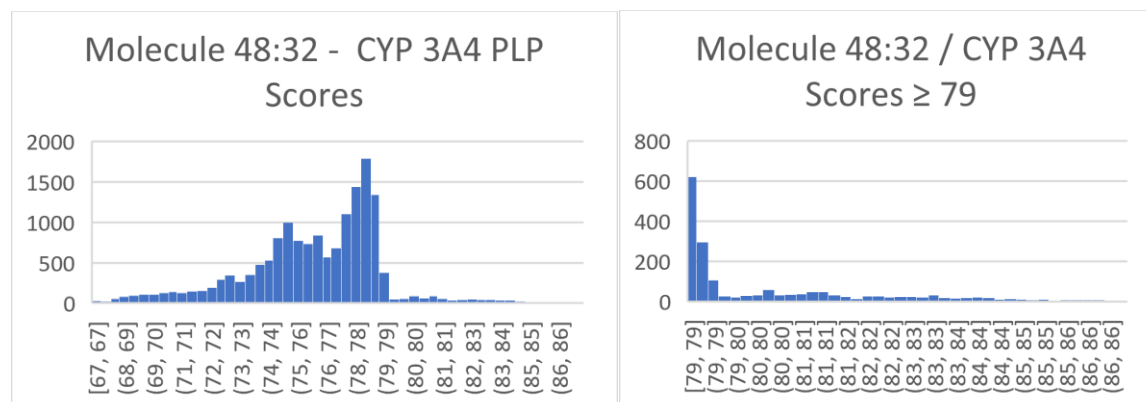


Figure 12: Histogram of GOLD docking job scores of Molecule 48:32 to CYP 3A4.



The combination of increased ligand interaction to the Mpro site by PLP 11.93, 17%, and in change in interaction to the two values 86.55 and 79 of CYP 3A4 by 2.11, 2%, or -5.44, 7%, is significant. In addition, the interaction differences between the 2 proteins goes from 13.76 to either 3.94 or -3.61; this is in activity difference of positive favor of Mpro to CYP 3A4 to positive in favor of Mpro. This is a change of Mpro-CYP 3A4 differences of 9.82 or 17.37, roughly 1.5 or 2.7 kCal.

The highest scoring docking information of 48:32 to Mpro and CYP 3A4 is

	Score	S(PLP)	S(hbond)	S(cho)	S(metal)	DE(clash)	DE(tors)	intcor
Mpro	82.61	-74.78	2.99	0	0	0	0.59	0.02
3A4	86.55	-81.65	1.00	0	.94	0	1.88	.02

There are 3 hydrogen bonds between molecule 48:32 and Mpro,

Donor molecule	Donor type	Donor heavy atom	Donor atom	Acceptor molecule	Acceptor atom	Score	Amino acid
L1	H	30	58	P1	1123	1.00	LEU 141
P1	H	1136	2470	L1	32	.99	GLY 143
P1	H	1312	3711	L1	40	1.00	GLU 166

one fewer than Nirmatrelvir to Mpro.

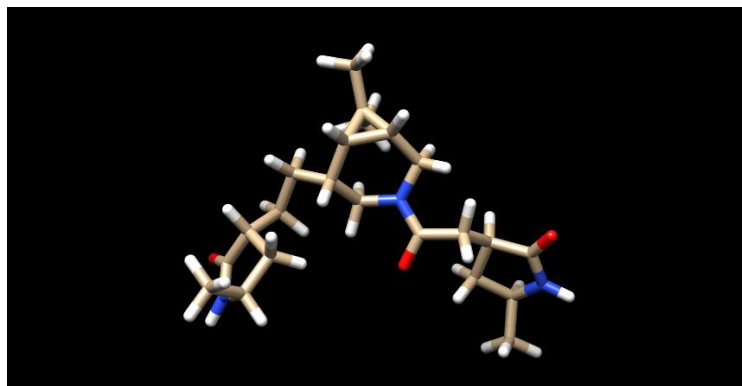
The overall increase of PLP 11.93 (17%) is roughly 1.8 kCal. This is a significant increase in protein-ligand interaction energy and could represent an increase of on-/off- binding of up to approximately x50 to x100.

Figure 13 shows the highest scoring pose of 48:32 to Mpro, the molecular diagram, and the docking to Mpro and CYP 3A4. In CCDC GOLD, the sidechains of some of the amino acids in the cavity are allowed to rotate to some extent. In Figure 14 an overlay of docked Nirmatrelvir to Mpro is shown with the protein from the docked 48:32 to Mpro. There is a small difference in the protein surface if the image is

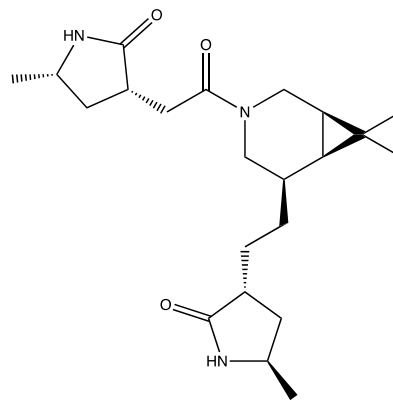
shown with the background protein of docked 48:32 to Mpro, but it is negligible. Figure 14 shows the superimposition of both in the CYP 3A4 enzyme cavity.

Figure 13: (a) Molecule 48:32, (b) chemdraw, (c) docked to Mpro, (d) docked to CYP 3A4

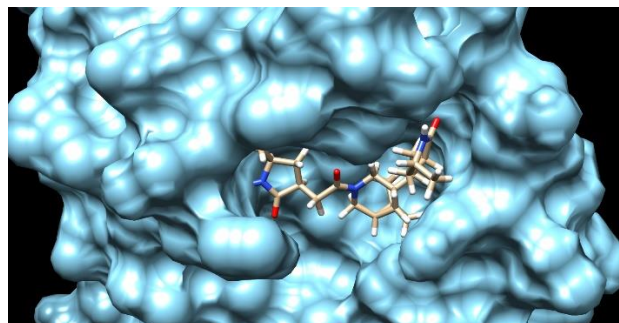
(a)



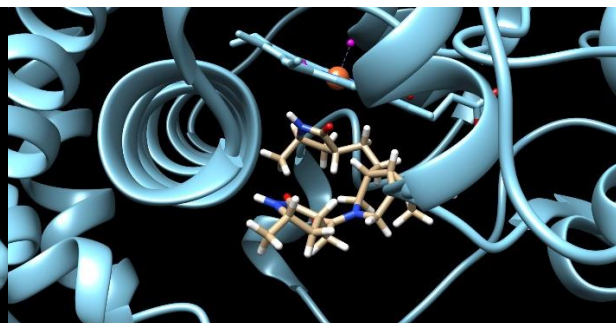
(b)



(c)

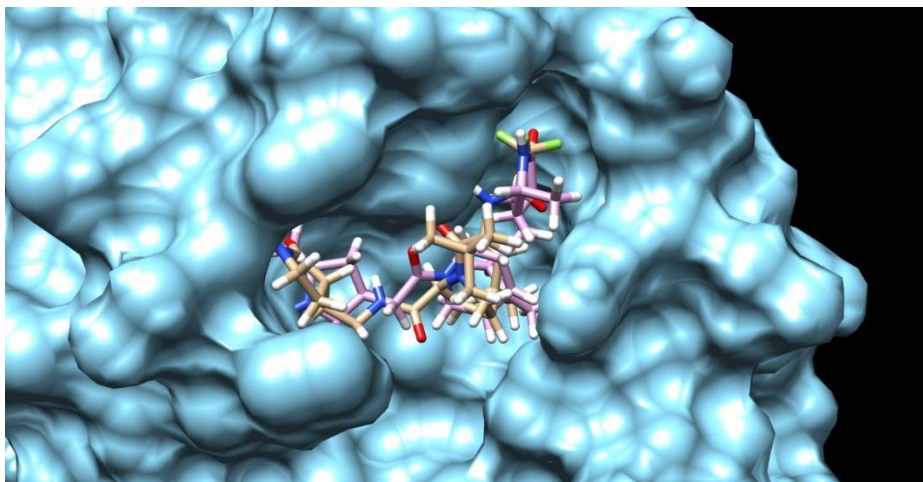


(d)



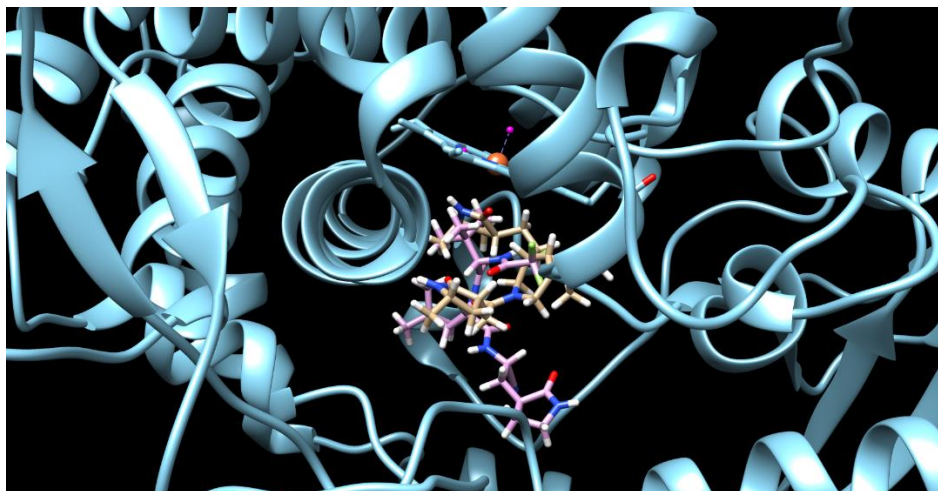
Molecule 48:32 in the X-shaped Mpro binding site spatially goes into 2 of the 3 caves and extends partly into the exiting surface tunnel. This is similar to Nirmatrelvir, which does the same, but 48:32 slightly farther. Its atomic content is different, and this will influence the interaction energy. It's 20% size reduction in this fitting make it more precise atomically as a ligand to target inhibitor. The per atom docking score of Molecule 48:32 is 2.95 as compared with to 2.02 of Nirmatrelvir; this is a 46% increase. It is strongly encouraged to open the mol2 files in a molecular viewer to see the different geometric aspects of binding of these molecules in overlay.

Figure 14: Docked Paxlovid and Molecule 48:32 on Mpro when Paxlovid is used in docking with the protein.



Both Nirmatrelvir and Molecule 48:32 bind into the CYP 3A4. In the Ligand GA software, the maximization of constructed molecules to one protein is only considered and the resulting set tested against the other, CYP 3A4.

Figure 15: Docked Paxlovid and Molecule 48:32 on CYP 3A4 when Paxlovid is used in docking with the protein.



Molecule 48:32 has an almost Z_2 or symmetry which is inherited from the same in the X-shaped cavity of the Mpro binding site. Were it not for an additional =O on one side it would be atomically reflective. This is in contrast to Nirmatrelvir which fills the legs of the cavity differently. The Ligand GA molecular evolution in finding optimal binders is guided by higher scoring molecules than the input population. The docking score per atom increased by 46% as a result, but it is interesting that the result points out symmetry of the Mpro cavity also.

In conclusion about Molecule 48:32 there are 2 more points. In regards to synthesis, it is a made of known fragments, 2 of which are in Nirmatrelvir. In biochemical activity it is small, 28 heavy atoms, that

can have atoms substituted to make the very complicated ADME process more appropriate. This work is computational, and there is no substitute for lab work and testing.

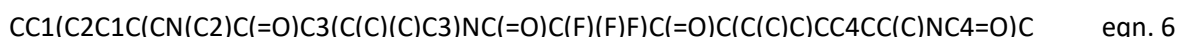
Analysis of Molecule 74:19

The discussion begins with the highest docking scores of Nirmatrelvir and Molecule 74:19 to Mpro and CYP 3A4,

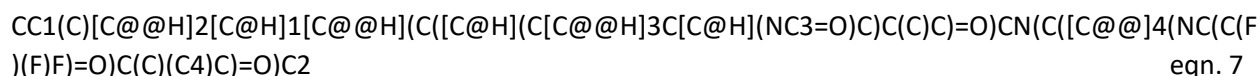
	Nirmatrelvir	Molecule 74:19
Mpro	70.68	91.11
CYP 3A4	84.44	94.27

This molecule fits in the set of 15922 as one of the highest scores to Mpro and average to CYP 3A4, $90.3+5.4=95.8$ is 1 sigma, Figure 9.

Molecule 74 stereoisomer 19 is in terms of non-isomeric SMILES,



and isomerically,



Its molecular properties relevant to ADME are:

Mass (Da)	Heavy atoms	Rotat. Bds	Hbond-don	Hbond-acc	ow
487	37	9	2	6	?

Molecule 74:19 is 98% the mass of Nirmatrelvir despite the increase of 2 atoms. It has 2 more heavy atoms and one less rotatable bonds than Nirmatrelvir. It has 4 rings not 3, 2 of which are 3-member C-rings. It is proportionally more carbon as it has fewer nitrogen's and oxygens, and is less massive. These parameters are seemingly better as far as tolerance and permittivity of it in the biological process of ingestion to the target protein and in ADME. Note that it has a 3-member C ring extending from it; as commented earlier on page 19, this can be edited and the docking re-examined.

The distribution of binding modes to Mpro is shown in Figure 16. There are several primary ones including a major center at 49 and one at 89. The one at 89 is much higher than the Nirmatrelvir at 68, a 32% increase. The histogram peak Figure 16(a) and resolved in (b) at scores 81 to 89 of Molecule 74:19 has many conformational states, suggesting conformational freedom to access the one at 91.11.

The score (interaction energy) distribution of docked states to CYP 3A4 is very broad with a center at 75. This distribution is in Figure 17. The binding cavity of CYP 3A4 is an open region that generic molecules can fit into, and although the maximum 74:19/CYP 3A4 is at 94.27, that is far from 75 and an outlier in the distribution.

Figure 16: Histogram of GOLD docking job scores of Molecule 74:19 to Mpro.

(a)

(b)

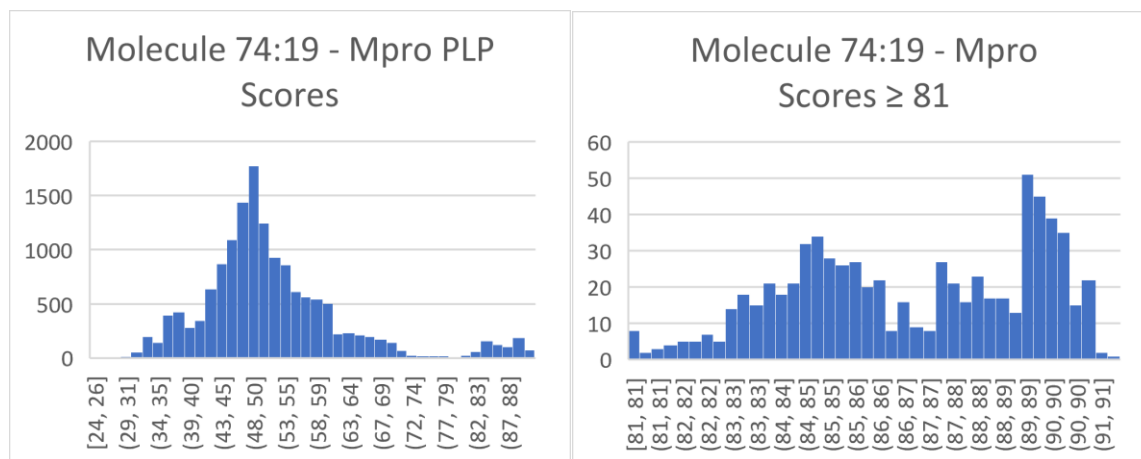
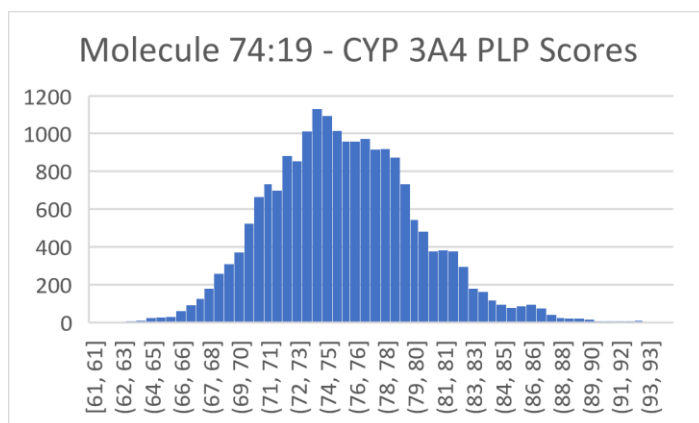


Figure 17: Histogram of GOLD docking job scores of Molecule 74:19 to CYP 3A4.



The increased ligand interaction to the Mpro site by the highest scores of Mpro and 74:19 is PLP 20.43, 29%. The highest pose to CYP 3A4 increases by 9.83, 12%. In addition, the interaction differences between the 2 proteins goes from 13.76 to 3.16; this is much more in favor of Mpro than to CYP 3A4 energetically. The point at 94.27 is an outlier in the distribution of scores and if a CYP 3A4 90 were used to compare the difference of scores from CYP 3A4 to Mpro would be negative. In addition to docking score uncertainty, modeling the molecule is comparable in interaction to Mpro as to CYP 3A4, unlike Nirmatrelvir with a separation of 13.76.

There are 6 hydrogen bonds in the highest scoring pose of 74:19/Mpro, 3 more than molecule 48:32/Mpro and 2 more than Nirmatrelvir/Mpro:

Donor molecule	Donor type	Donor heavy atom	Donor atom	Acceptor molecule	Acceptor atom	Score	Amino acid
L1	H	19	54	P1	1123	.99	LEU 141
L1	H	31	67	P1	1315	1.00	GLU 166
P1	H	1136	2470	L1	21	.17	GLY 143
P1	H	1140	2475	L1	21	.25	SER 144

P1	H	1146	2467	L1	21	.95	CYS 145
P1	H	1312	3711	L1	42	1.00	GLU 166

The docking information of 74:19 to Mpro and CYP 3A4 is

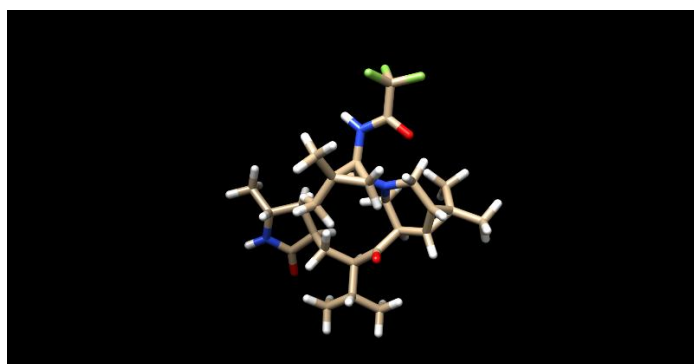
	Score	S(PLP)	S(hbond)	S(cho)	S(metal)	DE(clash)	DE(tors)	intcor
Mpro	91.11	-84.19	4.37	0	0	1.93	3.15	2.06
3A4	94.27	-84.74	2.00	0	1.00	0	1.95	1.43

The overall docking score of 74:19/Mpro is 91.11 and that of Nirmatrelvir/Mpro is 70.68, for an increase of PLP 20.43 or 29%. This is roughly 3.1 kCal (at approximately 1 kCal per 6.5 PLP score), meaning a fairly large increase in the ligand in-/off- states by up to a factor x1000, but this correspondence between kCal and population ratio x10 per kCal is not linear and a rough estimate.

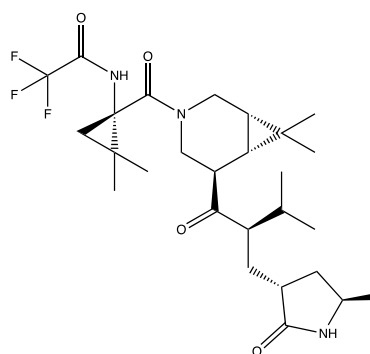
Figure 18 shows the highest scoring pose of 74:19 to Mpro, the molecular diagram, and the docking to Mpro and CYP 3A4. In Figure 5 an overlay of docked 74:19/Mpro and Nirmatrelvir is shown with the docked Nirmatrelvir/Mpro. Figure 6 shows the superimposition of both on the CYP 3A4 enzyme.

Figure 18: (a) Molecule 74:19, (b) ChemDraw, (c) docked to Mpro, (d) docked to CYP 3A4

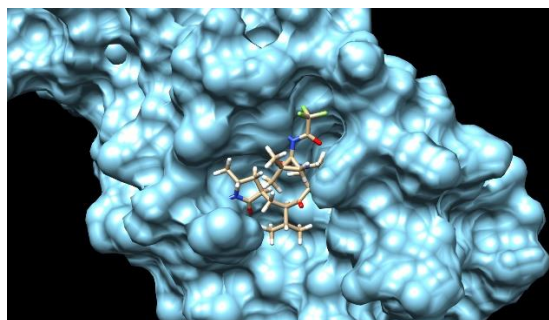
(a)



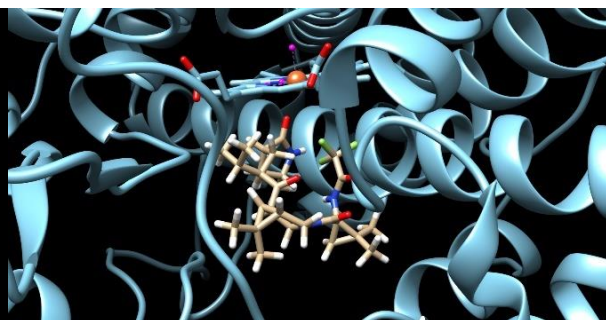
(b)



(c)



(d)



Molecule 74:19 in the X-shaped Mpro binding cavity extends into all 3 caves and partially extends into the surface tunnel. Spatially, it fills more of the cavity than Nirmatrelvir, which extends into 2 of the 3 caves and partly into the tunnel. The per atom docking score of 74:19 is 2.46 as compared with to 2.02 of Nirmatrelvir, 2.95 of Molecule 48:32. This is a 22% increase from Nirmatrelvir.

The overlays of molecule 74:19 and Nirmatrelvir attached to Mpro are shown in Figure 19. As before to illustrate the small differences in the rotatable protein sidechain that a ligand can make when attached these are shown with the Mpro docked with Nirmatrelvir and molecule 74:19. They are essentially the same. The overlays in the docking to CYP 3A4 are shown in Figure 20.

Figure 19: Overlay of docked Nirmatrelvir (tan) and Molecule 74:19 (pink) with the Nirmatrelvir/Mpro protein.

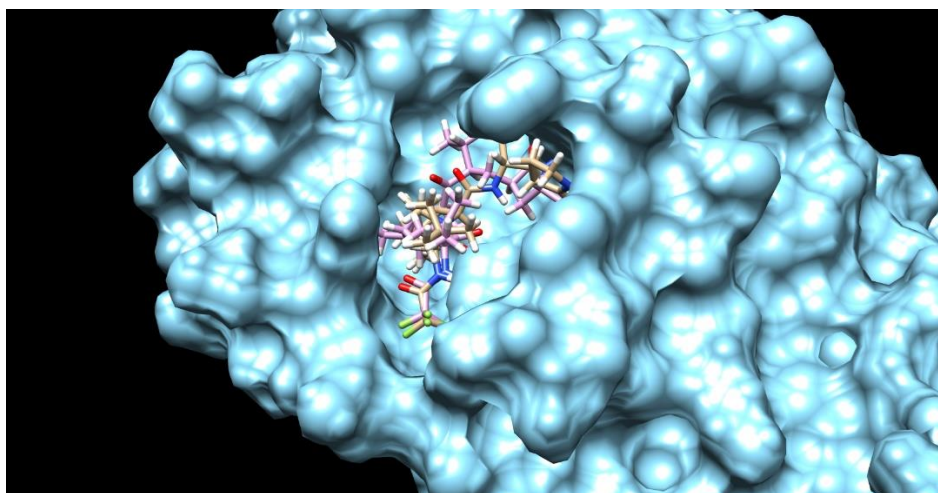
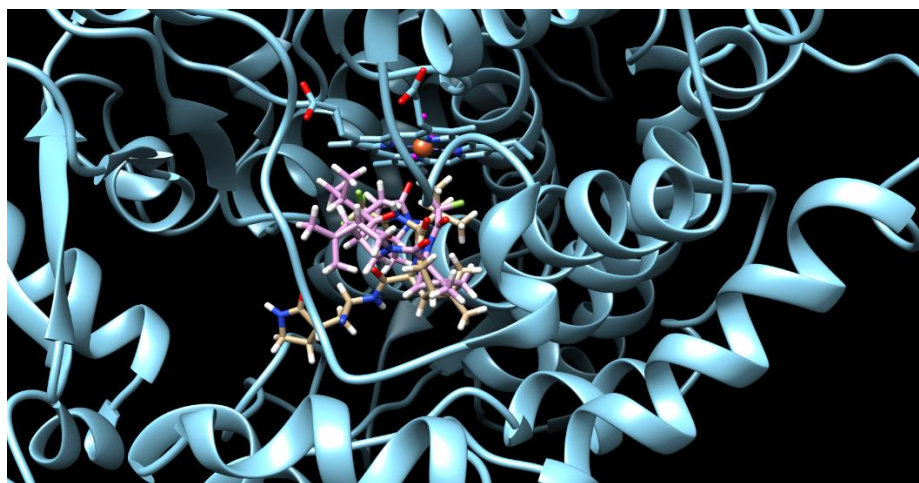


Figure 20: Overlay of docked Nirmatrelvir (tan) and Molecule 74:19 (pink) with Nirmatrelvir/CYP 3A4 protein.



Hydrogen bond contact analysis

The role of hydrogen bonding is relevant for both interaction and structural pinning of a ligand to a target protein. Biochemically, the availability of hydrogen bonding of a ligand can influence the transport of the small molecule through the bloodstream and barrier penetration through membranes, and permeability in different media [34]. In this section a comparison of the hydrogen bond contacts of Nirmatrelvir, Molecules 48:32 and 74:19 to Mpro. There is significant overlap in the set of hydrogen bonds. Individually the hydrogen bond information can be found in the gold docked mol2 files available in the figures directory of the supplementary information download.

Hydrogen bonds in the top scoring pose of the 3 molecules are listed in Table 3. The intersection of hydrogen bonding in the 3 molecules are highlighted. Green highlights the hydrogen bond contacts to GLY 143. Yellow shows the GLU 166 bonds. Blue shows the LEU 141. All of these amino residues have a shared bond with 2 or 3 of the ligands. The remaining 3, PHE 140, SER 144, and CYS 145 are unique to the different ligands. The amino acids are shown in Figure 21 in Mpro.

Table 4: Hydrogen bonds of Nirmatrelvir, molecule 48:32, and molecule 74:19.

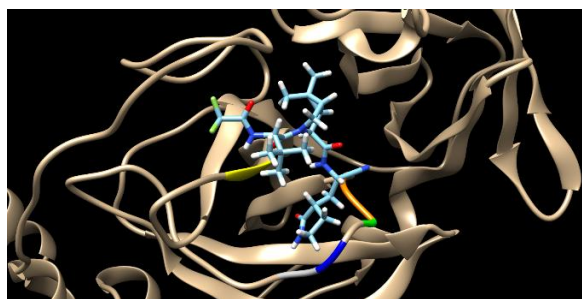
	Donor molecule	Donor type	Donor heavy atom	Donor atom	Acceptor molecule	Acceptor atom	Score	Amino acid
Nirmatrelvir	L1	H	19	56	P1	1315	1.00	GLU 166
	L1	H	36	64	P1	1112	.61	PHE 140
	P1	H	1136	2470	L1	40	.60	GLY 143
	P1	H	1312	3711	L1	12	.91	GLU 166
Mol 48:32	L1	H	30	58	P1	1123	1.00	LEU 141
	P1	H	1136	2470	L1	32	.99	GLY 143
	P1	H	1312	3711	L1	40	1.00	GLU 166
Mol 74:19	L1	H	19	54	P1	1123	.99	LEU 141
	L1	H	31	67	P1	1315	1.00	GLU 166
	P1	H	1136	2470	L1	21	.17	GLY 143
	P1	H	1140	2475	L1	21	.25	SER 144
	P1	H	1146	2467	L1	21	.95	CYS 145
	P1	H	1312	3711	L1	42	1.00	GLU 166

Geometrically the molecules fit into the X-shaped Mpro cavity differently. Nirmatrelvir and 48:32 share 2 of the X-legs and the surface tunnel. Molecule 74:19 occupies all 3 X-legs and the tunnel. This partly results in the overlaps of different hydrogen bonds of the 3 molecules and also explains the presence of 6 hydrogen bond contacts in molecule 74:16. None of the amino acids with these bonds are in a helix, but rather in strands.

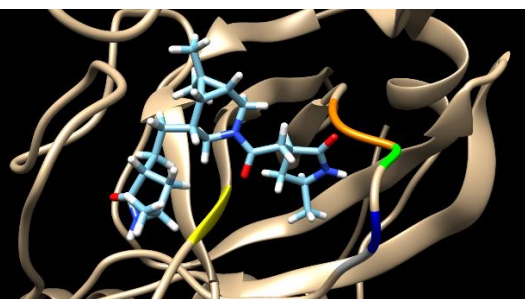
The shared pair of amino acids GLY 143 and GLU 166 by all 3 molecules are in different strands. The remaining amino acids LEU 141, PHE 140, SER 144, and CYS 145 are all adjacent and utilized differently in the 3 molecules in contacts.

Figure 21: Visual representation of highlighted amino acids in Table 4. Overlay with (a) Nirmatrelvir, (b) 48:32, (c) 74:19.

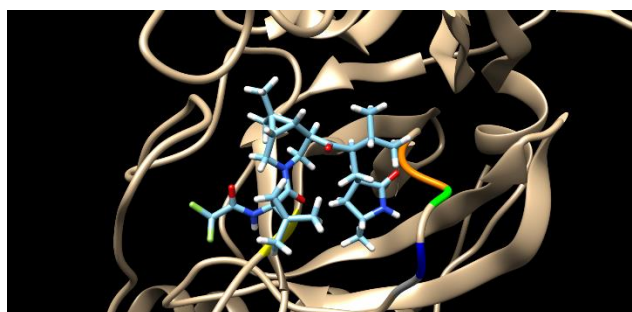
(a)



(b)



(c)



There is a significant similarity of the 3 networked hydrogen bonding in molecules. Note that in molecule 74:19 the oxygen acceptor atom 21 is shared in 3 different bonds totaling a 1.37 full hydrogen bond, and 2 of these are very weak. This atom is at the opposite end of the molecule than the trifluoromethyl group.

Section 5: Conclusions

Ligand GA, a computational molecular construction tool, is used to modify the Nirmatrelvir component of Paxlovid, an FDA EUA approved drug for treating SARS-Cov-2 post-infection to block viral replication. This work has 2 purposes: 1) demonstrate the utility of Ligand GA in a particular inhibitor problem, 2) produce realistic high binding small molecules to SARS-Cov-2 Mpro which are undetected in a usual repurposing or screening of known molecules or fragments.

Although metabolism is generally high within the liver enzyme CYP 3A4 and other CYPs of drug-like molecules, due presumably to the large featureless cavity, higher interacting small molecules can be designed that are more specific to the target, the viral replication activation site of Mpro of SARS-Cov-2. Up to 3 kCal of increased binding energy over dozens of previously proposed inhibitors of SARS-Cov-2 Mpro can be found in these Ligand GA designed molecules. This increase results in large binding activity, e.g. K_D , in a molecular on-/off- state of a protein-ligand complex.

A crucial goal is to eliminate the Ritonavir component of Paxlovid. The molecules generated from Ligand GA in the inhibitor/Mpro problem, which bind to the relevant set of amino acids of Mpro, are post-processed and some were selected due to their interactions both to the Mpro and CYP 3A4 enzyme. Two molecules are presented in detail and some others are mentioned and described.

Docking calculations are examined in random sampling to determine ligand conformational modes specific to the protein cavity, both in Mpro and CYP 3A4. Distributions of multiple docking run scores of

the same ligand can be used to determine the conformational modes of a small molecule attaching to a localized protein site. This in turn can be used with Boltzmann populations to determine contact probability with the protein site via a density of docking states. Some modes are more available for a molecule to occupy in its transition to binding and then in process to their highest Boltzmann weighted pose than to other modes.

The 2 example molecules 48:32 and 74:16 are plain looking molecules, with 48:32 almost reflective symmetric in the occupied Mpro region by Nirmatrelvir. The direct output only, without editing, of molecules from Ligand GA are presented and editing of molecules is expected in refinements of general Ligand GA output. 3 other molecules besides these 2 are also structurally described and several others mentioned. Information from the calculations are in the Supplementary Information.

This work is computational and no lab work has been done on these molecules, unlike with Nirmatrelvir and related molecules [13]. There are good similarities of the 2 proposed molecules with Nirmatrelvir structurally and in interaction; these molecules occupy the same location in the cavity as Nirmatrelvir and extend to other parts of it. The 2 molecules, and others in the presented set, appear ADME compatible, synthetically viable, editable without drastic change in docking, and soluble in different media as the other drug molecules in this work. Computationally, these are more Mpro target region specific than Nirmatrelvir, with increased binding, and low to mid-range binding to the liver enzyme CYP 3A4. There is a significant interaction energy increase and a gap reduction between Nirmatrelvir to Mpro versus CYP 3A4 from -2 kCal to zero.

Acknowledgements:

G.C. thanks Christian Heiss and James H. Prestegard for useful discussions. G.C. is grateful for the use of the Sapelo2 cluster at the Georgia Advanced Computing Resource Center and for the hospitality at the Complex Carbohydrate Research Center.

Statements and declarations

There are no conflicts of interest.

Supplementary information

Calculations, docking results files, figures, and the paper are located at <https://www.dropbox.com/sh/3vuza1s0gjmy1eo/AADLXPPe2Ve-N3RjMH3GoH5fa?dl=0>

Dropbox instead of GitHub is used because there is more space available for me in the Dropbox account. The Ligand GA software package is located at <https://github.com/GordonChalmers>

There are 3 directories: calculations, figures, and proteins.

The figures are made from mol2 files listed in READ_ME_figures.txt. The mol2 files are available in the calculations folder. For convenience the mol2 files used in the figures are also in a sub-directory mol2_files_fig in the figures directory. It is encouraged to look at the figures from the files in a molecular viewer.

The calculations folder contains docking results from the 2 steps in post-processing. These sub-directories can also be used in grepping for various reasons in molecular analysis. The files in using Ligand GA, .mat, prep, various other, are not in the download. These are available upon request, and

could be used to track evolution of individual molecules iteration by iteration (reconstruction of final molecules by process of mutation and crossover). These files also have the results non-isomeric molecules with scores < 79 which didn't screen into the 76 non-isomeric molecules used here.

References

1. Chalmers G, Ligand GA: A genetic algorithm for automated protein inhibitor design. In review at Journal of Computer Aided Drug Design, preprint. DOI:10.1101/2021.10.11.463970
2. MathWorks, Inc. MATLAB. (2020b). Retrieved from MathWorks: <https://www.mathworks.com/>
3. Weininger D, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, 28(1), 31-36. DOI:10.1021/ci00057a005
- Weininger D, Weininger A, Weininger J. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1998**, 29(2), 97-101. DOI:10.1021/ci00062/a008
- Weininger D. Smiles. 3. Depict. Graphical depiction of chemical structures. *J. Chem. Inf. Comput. Sci.* **1990**, 30(3), 237-243. DOI:10.1021/ci00067a005
4. Sadowski J, Gasteiger J., Klebe G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 1000-1008 DOI: 10.1021/ci00020a039
- Schwab CH. Conformations and 3D pharmacophore searching. *Drug Discovery Today: Technologies* 2010, 7(4), Winter **2010**, e245-e253 DOI: 10.1016/j.ddtec.2010.10.003
- Molecular Networks GmbH, Altamira, LLC. (2021). Corina. Retrieved from MN-AM Corina: <https://www.mn-am.com/products/corina>
- Schwab CH. Molecular Structure Representation in Chemoinformatics Applications, Schwab CH, BigChem Autumn 2017 School, Modena, Italy. Received from http://bigchem.eu/sites/default/files/School3_Schwab.pdf
5. Jones G, Willett P, Glen RC, Leach AR, Taylor R. (1997). Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, 267(3), 727-748. DOI: 10.1006/jmbi.1996.0897 PMID: 9126849
- Cambridge Crystallographic Data Centre. (2021). CCDC Discovery GOLD. Retrieved from GOLD Protein Ligand Docking Software: <https://www.ccdc.cam.ac.uk/solutions/csd-discovery/Components/Gold/>
6. Lipinski CA, Lombardo F, Dominy WB, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, 46(1-3), 3-26. DOI: 10.1016/s0169-409x(00)00129-0
- Lipinski C. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today Technologies* **2004**, 1(4), 337-341. PMID: 24981612 DOI: 10.1016/j.ddtec.2004.11.007
7. Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. *J. Comb. Chem.* **1999**, 1(1), 55-68. DOI: 10.1021/cc9800071.

8. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kobbie KD. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, 45(12), 2615-23. PMID: 12036371 DOI: 10.1021/jm020017n

9. Congreve M, Carr R, Murray C, Jhoti H. A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* **2003**, 8(19), 876-877. DOI: 10.1016/S1359-6446(03)02831-9

10. Benet LZ, Hosey CM, Ursu O, Oprea TI. BDDCS, the Rule of 5 and Drugability. *Adv Drug Deliv Rev.* 2016, 101, 89-98. DOI: 10.1016/j.addr.2016.05.007

11. Lucido MJ, Orlando BJ, Vecchio AJ, Malkowski MG. Crystal Structure of Aspirin-Acetylated Human Cyclooxygenase-2: Insight into the Formation of Products with Reversed Stereochemistry. *Biochemistry* **2016**, 55, 1226-1238. DOI: 10.1021/acs.biochem.5b01378 PMID: 26859324

Orlando BJ, Lucido MJ, Malkowski MG. The structure of Ibuprofen bound to cyclooxygenase-2. *J. Struct. Biol.* **2015**, 189, 62-66. DOI:10.1016/j.jsb.2014.11.005

12. Simeprevir. Retrieved from Pubchem: <https://pubchem.ncbi.nlm.nih.gov/compound/simeprevir>

Anson B, Mesecar A. (2020). X-ray structure of SARS-Cov-2 main protease bound to boceprevir at 1.45 Å. Published online. DOI:10.2210/pdb6WNP/pdb

13. Owen DA, et. al. An oral SARS-Cov-2 Mpro inhibitor clinical candidate for the treatment of Covid-19. *Science*, 24 December 2021, pp. 1586-1593. DOI: science.org/doi/10.1126/science.abl4784.

Making Paxlovid. Retrieved from: <https://www.science.org/content/blog-post/making-paxlovid>

14. Nirmatrelvir. Molecule PF-07321332. Retrieved from PubChem: <https://pubchem.ncbi.nlm.nih.gov/compound/155903259>

15. Ritonavir PDB ID 5VC0. Retrieved from: <https://www.rcsb.org/structure/5VC0>

Ritonavir. Retrieved from: <https://pubchem.ncbi.nlm.nih.gov/compound/392622>

16. PDBe-KB COVID-19 Data Portal at Protein Data Bank in Europe. (2021). Retrieved from: <https://www.ebi.ac.uk/pdbe/covid-19>

17. Matthew D Hall, James M Anderson, Annaliesa Anderson, David Baker, Jay Bradner, Kyle R Brimacombe, Elizabeth A Campbell, Kizzmekia S Corbett, Kara Carter, Sara Cherry, Lillian Chiang, Tomas Cihlar, Emmie de Wit, Mark Denison, Matthew Disney, Courtney V Fletcher, Stephanie L Ford-Scheimer, Matthias Götte, Abigail C Grossman, Frederick G Hayden, Daria J Hazuda, Charlotte A Lanteri, Hilary Marston, Andrew D Mesecar, Stephanie Moore, Jennifer O Nwankwo, Jules O'Rear, George Painter, Kumar Singh Saikatendu, Celia A Schiffer, Timothy P Sheahan, Pei-Yong Shi, Hugh D Smyth, Michael J Sofia, Marla Weetall, Sandra K Weller, Richard Whitley, Anthony S Fauci, Christopher P Austin, Francis S Collins, Anthony J Conley, Mindy I Davis. *The Journal of Infectious Diseases* **2021**, 224(S1)1:S1–S21. <https://doi.org/10.1093/infdis/jiab305>

18. Fischer W, et. al. Molnupiravir, an Oral Antiviral Treatment for COVID-19. DOI: <https://doi.org/10.1101/2021.06.17.21258639/>.

19. Molnupiravir. Retrieved from PubChem: <https://pubchem.ncbi.nlm.nih.gov/compound/145996610>.

20. Nirmatrelvir/Mpro complexes:

PDB ID 7SI9. Retrieved from the Protein Data Bank: <https://www.rcsb.org/structure/7SI9>

PDB ID 7VH8. Retrieved from the PDB: <https://www.rcsb.org/structure/7VH8>

PDB ID 7TE0. Retrieved from the PDB: <https://www.rcsb.org/structure/7TE0>

21. Nirmatrelvir predecessor PF-00835231. Retrieved from PubChem:

<https://pubchem.ncbi.nlm.nih.gov/compound/11561899>.

22. P450 (CYP) enzymes. Retrieved from:

[https://en.wikipedia.org/wiki/Cytochrome_P450#:~:text=Cytochromes%20P450%20\(CYPs\)%20are%20a,for%20hormone%20synthesis%20and%20breakdown](https://en.wikipedia.org/wiki/Cytochrome_P450#:~:text=Cytochromes%20P450%20(CYPs)%20are%20a,for%20hormone%20synthesis%20and%20breakdown).

23. Meunier B, de Visser SP, Shaik S. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.* 2004, 104, 9, 3947–3980. DOI: 10.1021.cr020443g.

24. McDonnell AM, Dang CH. Basic review of the cytochrome P450 system. *J. Adv. Pract. Oncol.* 2013 Jul-Aug; 4(4): 263–268. DOI: [10.6004/jadpro.2013.4.4.7](https://doi.org/10.6004/jadpro.2013.4.4.7) PMID: [25032007](https://pubmed.ncbi.nlm.nih.gov/25032007/) PMCID: PMC4093435

25. Drug interactions with Ritonavir. Retrieved from:

<https://www.covid19treatmentguidelines.nih.gov/therapies/statement-on-paxlovid-drug-drug-interactions/>

26. Ritonavir side effects. Retrieved from: <https://clinicalinfo.hiv.gov/en/drugs/ritonavir/patient>

Retrieved from: <https://medlineplus.gov/druginfo/meds/a696029>

27. Paxlovid FDA authorization drug sheet. Retrieved from:

<https://www.fda.gov/media/155050/download/>

28. Pettersen EF, Goddard GT, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, 25(13), 1605-PMID: 15264254 DOI: 10.1002/jcc.20084

University of California at San Francisco (UCSF) - Resource for Biocomputing, V. a. (Current). UCSF Chimera, an Extensible Molecular Modeling System. Retrieved from UCSF Chimera:

<https://www.cgl.ucsf.edu/chimera/>

29. Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **1996**, 14(1), 33-38. DOI: 10.1016/0263-7855(96)00018-5

University of Illinois at Urbana-Champaign, Theoretical and Computational Biophysics Group. (2021).

VMD - Visual Molecular Dynamics. Retrieved from VMD - Visual Molecular Dynamics:

<https://www.ks.uiuc.edu/Research/vmd/>

30. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. Retrieved from PyMOL:

<http://www.pymol.org/pymol>

31. Kingsley LJ, Brunet V, Lelais G, McCloskey S, Milliken K, Leija E, Fuhs SR, Wang K, Zhou E, Spraggon G. (2019). Development of a virtual reality platform for effective communication of structural data in drug

discovery. *Journal of Molecular Graphics and Modeling* **2019**, 89, 234-242.
DOI:10.1016/j.jmgm.2019.03.010

Nanome, Inc. (2021). Nanome. Retrieved from Nanome: <https://nanome.ai/>

32. Møllendal H. A Microwave and Quantum Chemical Study of (Trifluoromethyl)thiolacetic Acid, CF₃COSH, a Compound with an Unusual Double-Minimum Potential. *J. Phys. Chem. A* **2007**, 111, 1891-1898. DOI: 10.1021/jp0677290 PMCID: PMC7486594 PMID: 32918656

33. Bafna K, Krug RM, Montelione GT. (2020). Structural Similarity of SARS-CoV2 M pro and HCV NS3/4A Proteases Suggests New Approaches for Identifying Existing Drugs Useful as COVID-19 Therapeutics. DOI: 10.26434/chemrxiv.12153615. Preprint. PMID: 32511291 PMCID: PMC7263768

34. Bafna K, White K, Harish B, Rosales R, Ramelot TA, Acton TB, Moreno E, Kehrer T, Miorin L, Royer CA, García-Sastre A, Krug RM, Montelione GT. Hepatitis C virus drugs that inhibit SARS-CoV-2 papain-like protease synergize with remdesivir to suppress viral replication in cell culture. *Cell Rep.* **2021**, 35(7): 109133. DOI: 10.1016/j.celrep.2021.109133, PMID: 33984267, PMCID: PMC8075848

35. Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, Ruan L, Song B, Cai Y, Wei M, Li X, Xia J, Chen N, Xiang J, Yu T, Bai T, Xie X, Zhang L, Li C, Yuan Y, Chen H, Li H, Huang H, Tu S, Gong F, Liu Y, Wei Y, Dong C, Zhou F, Gu X, Xu J, Liu Z, Zhang Y, Li H, Shang L, Wang K, Li K, Zhou X, Dong X, Qu Z, Lu S, Hu X, Ruan S, Luo S, Wu J, Peng L, Cheng F, Pan L, Zou J, Jia C, Wang J, Liu X, Wang S, Wu X, Ge Q, He J, Zhan H, Qiu F, Guo L, Huang C, Jaki T, Hayden FG, Horby PW, Zhang D, Wang . A trial of lopinavir-ritonavir in adults hospitalized with severe Covid-19. *N. Engl. J. Med.* **2020** doi: 10.1056/NEJMoa2001282.

36. Computing resources. <https://www.gacrc.uga.edu>,
https://wiki.gacrc.uga.edu/wiki/Georgia_Advanced_Computing_Resource_Center/

37. Machine Learning for Pharmaceutical Discovery and Synthesis Consortium (MLPDS)
<https://mlpds.mit.edu>, on-line synthesis prediction tool <https://askcos.mit.edu>