Exploring Fragment-Based Methods for Designing Inhibitors of SARS-CoV-2 NSP14/NSP10

SN: 20067984
Correspondence: Department of Pharmaceutical & Biological Chemistry, UCL School of Pharmacy, 29-39 Brunswick Square, WC1N 1AX

ABSTRACT: SARS-CoV-2's emergence has presented a serious risk, and Covid-19 has affected the health of individuals, society, and economy. Although vaccines have been developed quickly, it has been difficult to repurpose current medications for efficient treatment. Novel medications that target SARS-CoV-2 and other coronaviruses are needed for future pandemics. SARS-CoV-2 non-structural protein 10 (NSP10) and NSP14 are known to play a crucial role in viral replication and transcription. Since these enzymes exhibit exoribonuclease and methyltransferase activities, interfering with RNA proofreading and cap formation may be a useful tactic to stop viral replication.
The aim of the research project is to identify a compound capable of strongly binding and inhibiting the action of NSP10/NSP14 protein. Our objectives also include designing new high molecular weight molecules, analysing the new fragment data, and determining the effect of the NSP10/NSP14 protein on the virus.
In this research, we examined eighteen fragments that are presently accessible on Fragalysis, a software programme made for analysing the three-dimensional structures of molecules.
Our plan involves analysing these 18 fragments, categorizing them into similar functional groups, and identifying overlapping areas. This process will enable us to design new high-molecular-weight molecules with enhanced binding affinity to the protein by using optimization strategies such as merging, linking, and growing.
The Lipinski rule was employed as a measure of the molecule's efficacy; however, it is essential to note that the feasibility of synthesising these molecules may pose multiple challenges. Further investigations and efforts will also need to be undertaken to test the efficacy of these inhibitors once they are synthesized through lab testing and obtaining physical data, such as IC50.

1. INTRODUCTION

The COVID-19 pandemic, caused by the highly transmissible airborne virus known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1], has had a profound global impact.
SARS-CoV-2 is a positive-sense single-stranded RNA virus with a complex genome encoding 29 proteins, including four structural proteins.[2] SARS-CoV-2's genome is approximately 29.9 kilobases in length and encompasses 14 open reading frames (ORFs). The first two ORFs, ORF1a and ORF1b, give rise to two large polyproteins, pp1a and pp1ab, which are subsequently cleaved into 16 non-structural proteins essential for the virus's replication and transcription.[3]
To evade the host cell's immune system and manipulate the translation apparatus, the viral RNA must develop a cap structure at its 5' end, similar to eukaryotic mRNAs.[3] Interestingly, coronaviruses employ a multi-step process to mimic the host cell's capping machinery installation. The SARS-CoV-2 virus binds with high affinity to the human ACE2 receptor to enter host cells and replicates its RNA genome.[4]
While multiple vaccines and treatments have been developed to combat the pandemic, research endeavors continue to search for therapeutic agents targeting the virus, its variants, or potential future zoonotic coronaviruses.[5] Notably, one antiviral treatment for COVID-19 is remdesivir, though its efficacy has been a subject of debate.[6] Additionally, non-structural proteins like NSP10 and NSP14, which play crucial roles in viral replication and RNA capping, are among the potential targets for combatting SARS-CoV-2.[3] The genomic complexity and unique mechanisms of this virus make it a significant focus of ongoing research and public health efforts.

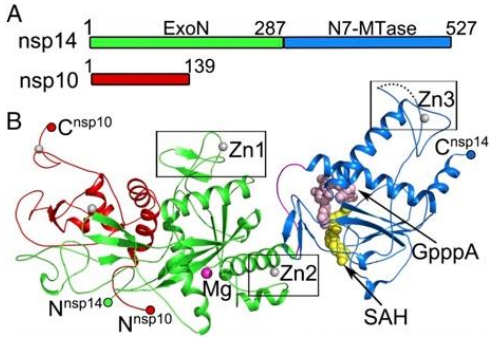


Figure 1. shows the overall structure of the nsp14–nsp10 complex. (A) represents the domain organization of nsp14 and nsp10. (B) represents the structure of the nsp14–nsp10 heterodimer. Nsp10, the ExoN domain, the N7-MTase domain, and the loop at the N terminus of N7-MTase are coloured in red, green, marine, and pink, respectively.[7]

Non-structural protein 14 (NSP14) of SARS-CoV-2, serves a crucial role in the virus's replication and transcription.[7] NSP14 is a bifunctional enzyme with distinct activities at its N- and C-termini.
At its N-terminus, it possesses a 3'-5' exoribonuclease (ExoN) activity that functions as a proof-reader, reducing the mutation rate of the virus

and preventing excessive genetic variability in its unusually large RNA genome.[8] This proofreading activity helps maintain the integrity of the virus's genetic material.[7] NSP14's exonuclease activity also plays a role in reducing the host innate antiviral immune response. When SARS-CoV-2 replicates, it produces double-stranded RNA (dsRNA) intermediates, which are known to cause type I interferon (IFN-I) to be released in the host. In SARS-CoV-2, NSP14 has been found to be a potent interferon antagonist. This means that NSP14's Exonuclease activity allows it to cleave dsRNA, indicating that it may be involved in breaking down viral dsRNA replication intermediates, thereby inhibiting the host's innate immune response from activating and regulating viral genome recombination.[6, 9]

However, the exonuclease activity of NSP14 is stimulated by its cofactor, NSP10, and is only fully unleashed in its presence.[7] Furthermore, the C-terminal domain of NSP14 functions as a guanine-N7 methyltransferase (N7-MTase) for mRNA capping, which is crucial for viral genome replication and evasion of host defenses.[7] MTases are responsible for methylating a variety of substrates (such as proteins and nucleic acids),[10] this is crucial as Viral RNA cap synthesis involves two methylation steps. Additionally, it has been demonstrated that NSP14 MTase plays a crucial role in the life cycle of coronaviruses, affecting their replication efficiency and significantly imparing the immune system's ability to respond to viral infection.[3] NSP14, with its multifaceted functions, has become an attractive target for medicinal chemistry research to develop antiviral therapies against SARS-CoV-2, complementing traditional targets like the RNA-dependent RNA polymerase or viral proteases.[3] Understanding the molecular basis of nsp14's activities is essential for advancing our knowledge of coronavirus replication and for the development of potential antiviral treatments.[7]

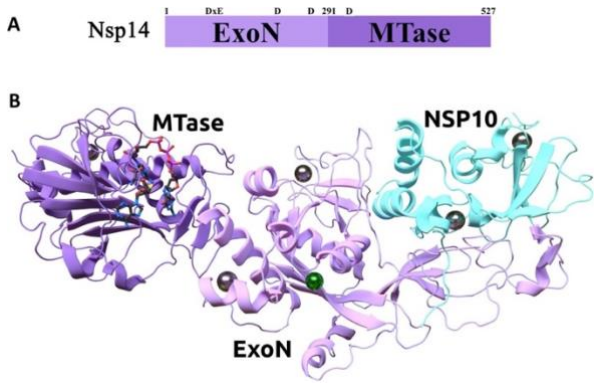


Figure 2. (A) shows nsp14 domain organization. ExoN domain; MTase domain. (B) shows the nsp10-nsp14 complex of SARS-CoV-2. The nsp10 is coloured in cyan. The MTase and ExoN domains of nsp14 are coloured in dark and light purple, respectively.[9]

Non-structural protein 10 (nsp10) is a pivotal player in the SARS-CoV-2 viral replication machinery, forming a complex with two other viral non-structural proteins, NSP14 and nsp16.[7] NSP10 binds to the N-terminus of NSP14, activating its 3'-5' exoribonuclease (ExoN) activity, a function that aids in proofreading and maintaining the integrity of the virus's unusually large RNA genome. While NSP10 enhances the ExoN activity of NSP14, it does not appear to be required for the stimulation of the N7-methyltransferase (N7-MTase) function.[8] Disruption of the NSP10-NSP14 interaction has been shown to abrogate the activation of NSP14's ExoN activity.[11] Additionally, NSP10's role in regulating viral enzymatic activities and replication-transcription complex assembly makes it an attractive target for drug design programs.[11] Identifying fragments that bind to SARS-CoV-2 NSP10 is a crucial step in structure-based drug design, opening the door to potential antiviral therapies. Additionally, studies have revealed that NSP14 and 10 have a weak interaction, suggesting that small molecules may interfere with the formation of the NSP10-NSP14 complexes.[8] Understanding the molecular interactions within the NSP10-NSP14 complex is essential for advancing our knowledge of SARS-CoV-2 replication and for developing strategies to combat the virus effectively.

In this thesis, we plan to identify small molecules that can bind and prevent NSP10 and NSP14 from associating as a way of suppressing viral replication. Targeting NSP10, and consequently indirectly NSP14, offers a unique and selective method to stop SARS-CoV-2 replication because NSP10 is not present in host cells.[8] By targeting NSP10 and NSP14, researchers have discovered promising, potent inhibitors.

1.1 LITERATURE INHIBITORS: ENZYME INHIBITORS OF NSP10/14

Compound S-adenosyl-L-homocysteine (SAH) is an example of an inhibitor (that is commonly used as a reference or control compound to test the inhibitory effect of other compounds); it demonstrated significant inhibitory activity against the NSP14 methyltransferase activities, with IC_{50} values of 0.22 μM.[5]

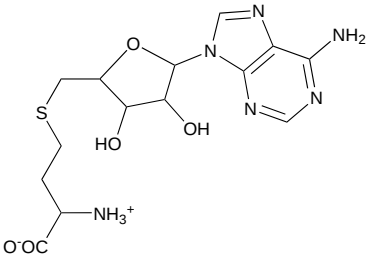


Figure 3.
Shows the structure of S-adenosyl-L-homocysteine (SAH)

In addition, researchers have discovered binding sites for fragments on NSP10. One such discovery involves a fragment that is able to bind to two ligand binding sites on NSP10. This fragment is known as VT00022 (as shown in figure 5), which may be able to interfere with the interaction between NSP14 and

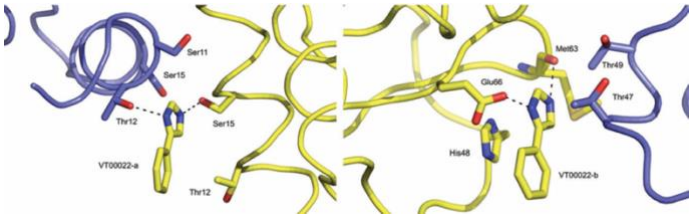


Figure 4. ligand binding sites of nsp10 for fragment VT00022. The protein is shown in yellow ribbon. Hydrogen bonds are indicated as black dashed lines. The image of the shows VT00022 binding in the nsp14–nsp10 interfaces and the right-hand side

NSP10 because its binding site is situated at the interface between NSP14 and NSP10.[8] At the interface between NSP10 and its second binding partner, NSP16, another ligand binding site was found. These fragments may obstruct the NSP14-NSP10 complex's formation, offering a possible target for antiviral therapy.[8] This fragment was identified using X-ray based fragment screening (XFS) and thermal shift assay (TSA). The affinities of the fragment was quantified using Microscale thermophoresis (MST) experiments.

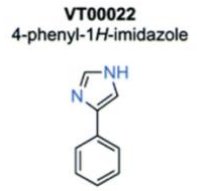


Figure 5. Structure of VT00022.[8]

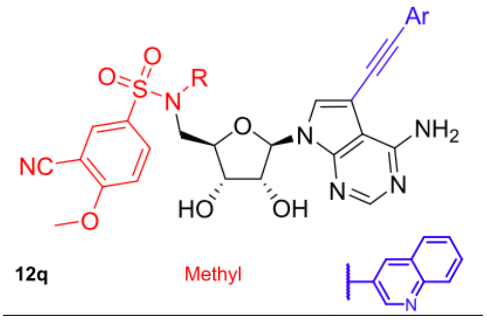


Figure 6. Structure of 12q. R represents methyl group.[3]

A study has discovered a compound known as compound 12q (shown in the figure) that is capable of binding to NSP14 methyltransferase. Compound 12q is synthesised by modifying the sulfonamide N-alkyl substituents on SAH derivatives.[3] It demonstrated an IC_{50} of 19 nM, marking it as one of the most potent NSP14 inhibitors ever documented.[3] This compound demonstrated low toxicity *in vitro* and do not inhibit tested human MTases. Furthermore, they are the first inhibitors with such strong inhibitory activity against SARS-CoV-2 NSP14 that can easily enter cells.[3] However, their limited effectiveness in inhibiting SARS-CoV-2 replication *in vitro* highlights the need for further investigation to fully comprehend the enzyme's role in viral replication in cell lines.[3]

Another study has identified a compound called patulin and aurintricarboxylic acid (ATA), which are

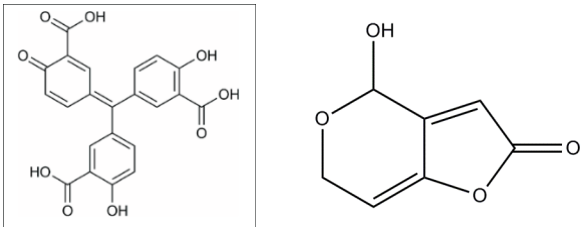


Figure 7. Structure of ATA on the left-hand side and structure of Patulin on the right-hand side[6]

in vitro inhibitors that target the NSP14/NSP10 ExoN enzyme.[6] Patulin showed an IC_{50} of 1.8 mM and ATA showed an IC_{50} of 10.3 mM.[6] They were found to reduce the proliferation of SARS-CoV-2 in cell-based assays.[6] However, further research is needed to determine whether these inhibitors directly target NSP14/NSP10 ExoN.[6] The assays were conducted using VERO E6 cells, (this is a common Green Monkey Chlorocebus sabaeus cell line), and future studies may investigate the inhibitors' efficacy in human cell models of SARS-CoV-2 infection.[6]

Bi(III)-based compounds (such as RBC, $Bi(NAC)_3$, and 4SABi) were found to allosterically inhibit both the ExoN and MTase activities of SARS-CoV-2 NSP14/10.[12] RBC, $Bi(NAC)_3$, and 4SABi demonstrated IC_{50} values of 1.03 ± 0.24, 1.33 ± 0.26, and 1.36 ± 0.28 μM for its MTase activity and 0.53 ± 0.07, 0.80 ± 0.06, and 0.84 ± 0.04 μM for its ExoN activity, respectively, revealing high potentials of these compounds to be developed as anti-SARS-CoV-2 drugs.[12] This inhibition occurred because these compounds displaced Zn(II) ions from the zinc-finger sites of the enzymes by binding with Bi(III)

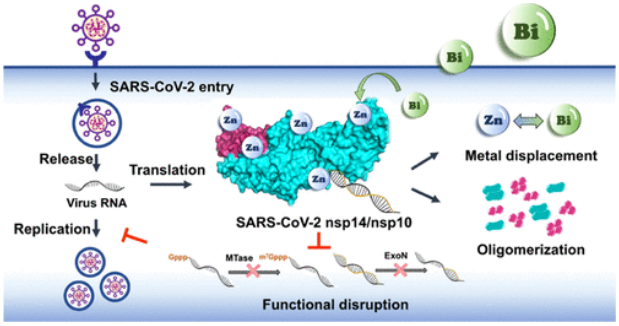


Figure 8. demonstrates Bi(III) binding to both nsp14 and nsp10, resulting in the release of Zn(II) ions and disrupting the quaternary structure of the protein.[12]

ions.[12] Additionally, the binding of Bi(III) to NSP10 and NSP14 led to changes in the quaternary structures of these enzymes. Furthermore, the antiviral properties of Bi(III)-based NSP14/NSP10 inhibitors were effective against various SARS-CoV-2 variants in mammalian cells.[13]

This project is a Type C Data-Analysis with a word count of 5624 words formatted in the printed style of Journal of Medicinal Chemistry

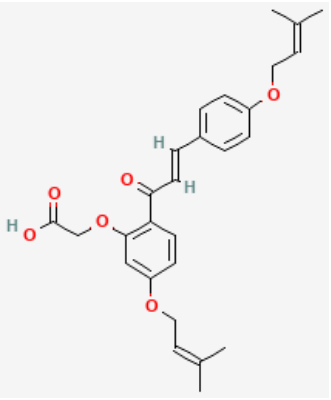


Figure 9. Structure of Sofalcone.[13]

Sofalcone was identified as an NSP14-NSP10 inhibitor that increases remdesivir's potency.[13] A model compound, sofalcone, was shown to inhibit the exonuclease activity of SARS-CoV-2 and synergistically enhances the antiviral effect of remdesivir, preventing the replication of SARS-CoV-2.[13] Sofalcone showed high synergistic effect with remdesivir, lowering the EC50 values of remdesivir by ~5-fold.

A different study has concluded that the best, potentially clinically useful NSP14–NSP10 inhibitor identified is the Mpro inhibitor ebselen, where the selenium moiety plays an important role mediating its inhibitory effects.[14] Ebselen demonstrated an IC_{50} values of 3.3 ± 0.09 uM.[14] The exact mechanism of how ebselen inhibits NSP14-NSP10 remains unclear, but it may involve selenium's role in coordinating or displacing 2 zinc atoms within these proteins. Regardless of the specific mechanisms, ebselen's ability to target two crucial enzymes needed for SARS-CoV-2 replication suggests that further research is needed to evaluate its potential use in combating COVID-19, either on its own or in combination with other agents.

Four potential inhibitors of NSP14 were identified, which also exhibited antiviral effectiveness in a cell-based SARS-CoV-2 infection model.[15] All four compounds displayed antiviral activity with limited cytotoxicity. The most effective inhibitors were PF-03882845 (EC50 = 10.97 µM), Inauhzin (EC50 = 12.96 µM), and Trifluperidol (EC50 = 14.9 µM).[15] Lomeguatrib exhibited lower effectiveness against viral replication, with an EC50 = 59.84 µM, and showed slight cytotoxicity at effective concentrations.[16] This highlights the antiviral potential of the identified compounds in mammalian cells with minimal adverse effects.

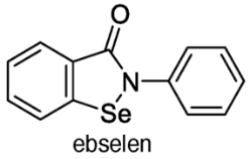


Figure 10. Structure of ebselen.[14]

Modified SAH molecules were designed as inhibitors targeting the NSP14 MTase enzyme by attaching aromatic substituents to position 7 using appropriate linkers.[16] These compounds demonstrated potent inhibitory activity against NSP14, with IC_{50} values ranging from 3 to 37 nM. Among them, compound 16 (TO507) was the most effective, with an IC_{50} of 3±0.5 nM.

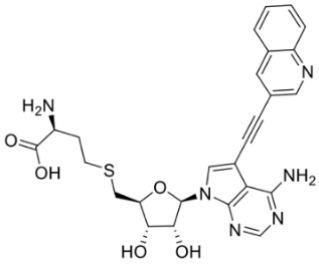


Figure 11. Structure of compound 16 (TO507).[16]

2. BACKGROUND

There are challenges of targeting the NSP10/14 pair due to protein-protein interaction. The NSP10/14 complex presents one of the most challenging targets in the field of drug discovery, primarily due to its nature as a protein-protein interaction (PPI) target. Several factors contribute to the complexity and difficulty associated with PPI targets:

1. In PPIs, proteins often interact over large, flat surfaces, making it hard to create small molecules that can effectively disrupt or modify these interactions. Traditional small-molecule drugs might find it difficult to cover such extensive binding areas. Moreover, it becomes challenging to improve a drug's effectiveness once the binding site moves past well-established "hot spots" (specific interaction points), especially for PPIs.[17]
2. PPIs binding sites might be shallow and lack well-defined pockets, making it easier for small molecules to target them effectively. This makes it difficult to find compounds that can bind strongly and specifically to these sites. Protein surfaces contain small grooves, pockets, and indentations but are not necessarily always flat. The contact surfaces of a protein-protein complex are defined by these characteristics on the protein's surface, which can hold water molecules and pieces of a partner protein in the bound state.[18]
3. PPIs frequently involve dynamic, flexible interfaces, which makes it challenging to create molecules that can both inhibit and stably bind to the interaction. The innate flexibility of these interfaces makes it difficult to create compounds that can successfully hinder protein-protein binding. Fortunately, a great deal of progress has been made in comprehending these interfaces' dynamic behaviour. High-resolution structure prediction and molecular design advances are expected to have a major impact on biology and medicine, including the understanding and targeting of PPIs. Despite the challenges mentioned, certain characteristics of proteins, such as hot spots and allosteric sites, may be efficiently targeted by small molecules.[18]
4. It can be difficult to achieve specificity when attempting to break a specific protein-protein interaction (PPI). Specificity means targeting only the intended PPI without affecting other essential interactions or cellular functions. The challenge is avoiding accidental interactions with various proteins, which can cause negative side effects and reduce a medication's efficacy. Natural products and compounds derived from them have an advantage because of the biological validation of their chemical structures, which may result in high specificity and binding affinity to a variety of biological targets.[18]
5. disrupting a PPI may lead to significant alterations in cellular pathways and unexpected outcomes. Proteins can selectively interact and form complexes that influence biological pathways.[18]

3. AIMS

- Conduct a comprehensive analysis of the fragment data, with a specific focus on observing the interactions between the NSP10/14 protein and the identified fragments.
- Design new high molecular weight molecules that are able to bind to NSP14/10 with high affinity and prevent the NSP14/10 complex from associating
- To share the designs of the new molecules online, facilitating a collaborative environment wherein fellow researchers can actively engage and contribute to the ongoing advancements in this molecular design

In this research, our specific focus lies in investigating fragments that bind to the interface between NSP14 and NSP10. Currently, approximately 18 fragments are listed on Fragalysis, demonstrating binding at the interface site. Our plan involves analysing these fragments, categorizing them into similar functional groups, and identifying overlapping areas. This process will enable us to design new high-molecular-weight molecules with enhanced binding affinity to the protein.

4. METHOD

The project began when Joseph A Newman discovered fragments that bind to the protein pair SARS-CoV-2 Nsp10/14, and the fragment he discovered had binding affinity for both proteins. They determined the structure of NSP14 in complex with the cap analogue ^{7Me}GpppG (a molecule that is designed to micmic the structure of the 5'cap of RNA) and observed conformational changes within a SAM/SAH interacting loop that plays a critical role in viral mRNA capping, providing new insights into MTase activity. They ran an X-ray fragment screen on NSP14 and found 72 hits that corresponded to inhibition sites in the ExoN and MTase domains. These fragments are excellent starting points for the structure-guided development of NSP14 inhibitors.[19]

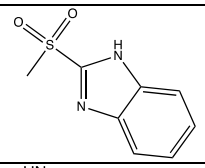
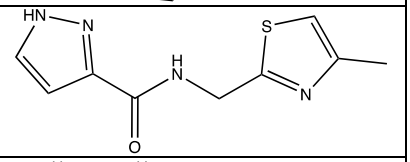
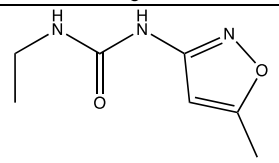
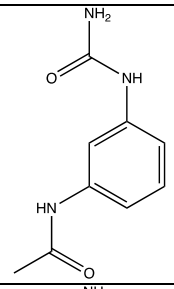
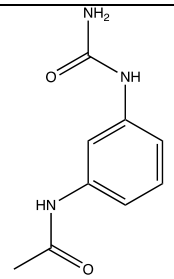
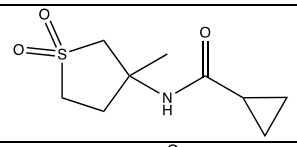
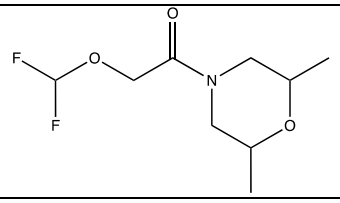
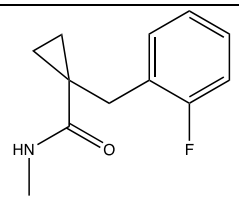
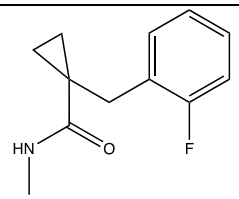
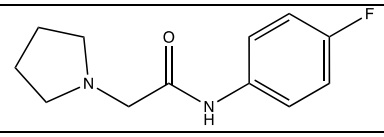
Our objective is to identify fragments capable of strongly binding and inhibiting the action of NSP10/NSP14 protein. We have chosen to analyse 18 fragments that are currently available on Fragalysis, a software designed for the three-dimensional structure analysis of various molecules. We chose to investigate these 18 fragments because they bind at the interface between NSP10 and NSP14. We are focusing on fragments with a binding site at the NSP10/NSP14 interface because we believe they will disrupt the formation of the NSP10-NSP14 complex and prevent the proteins from associating.

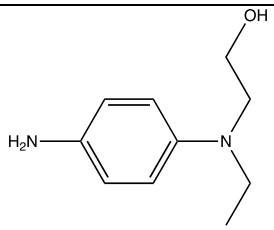
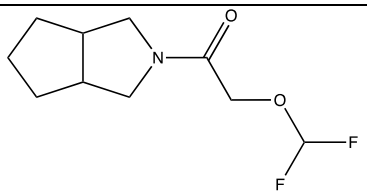
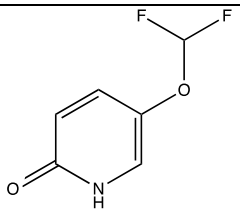
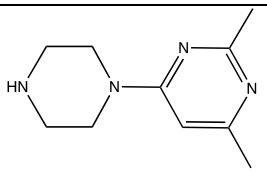
4.1 Fragalysis

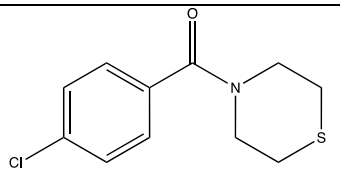
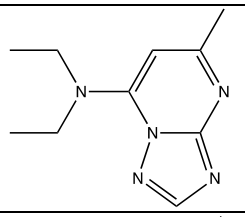
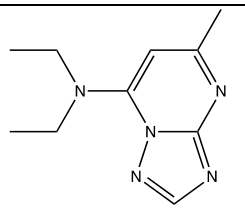
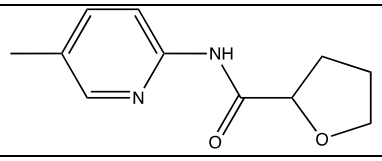
Fragalysis is a web-based platform that was initially inspired by the fragment screening experiment carried out at Diamond, a synchrotron light source facility. Fragalysis is essentially an open-source platform that aims to make the progression from fragment hits to more potent binders faster and more cost-effective.[20] With an intuitive web-based interface, it is suitable for non-expert users outside of the field of computational medicinal chemistry as well. Because Fragalysis curates and presents the generated data in an easily readable manner, it also makes data sharing with the larger scientific community easier. In the field of fragment-based drug discovery, the platform is an effort to bring together experimental and computational methods.

Chemical structure of the 18 Fragments

The table below show the chemical structure of all the 18 fragments (listed from Fragalysis) that binds to NSP14 at the interface between NSP14 and NSP10

Fragment	Chemical structure
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

11	
12	
13	
14	

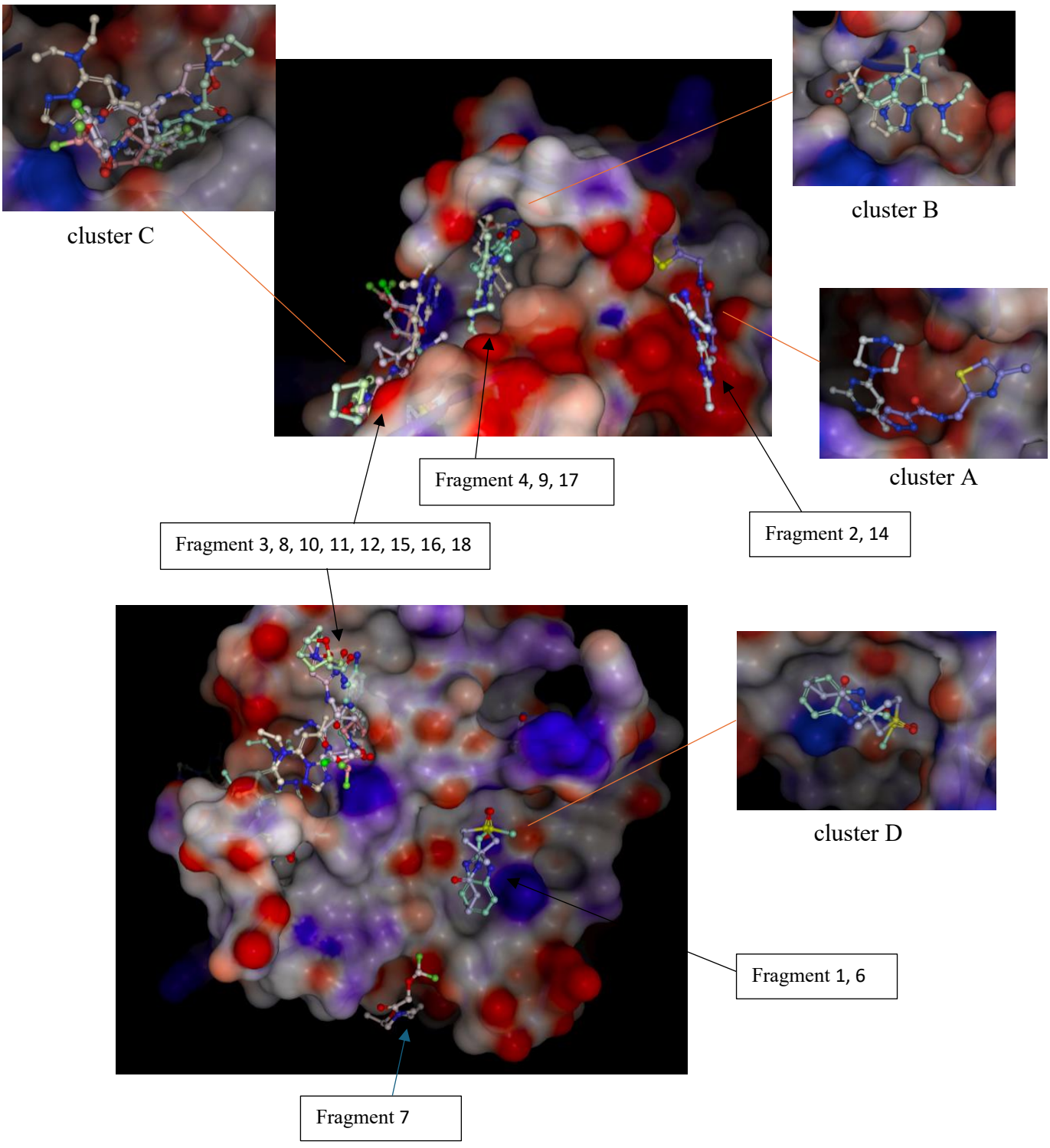
15	
16	
17	
18	

It has been observed that fragment 4 and 5 share identical chemical structures but exhibit binding at distinct sites on the NSP14 enzyme. This parallel phenomenon is similarly observed with fragment pairs 8 and 9, as well as 16 and 17.

4.2 Analysing the 18 Fragments

The process was initiated by analysing 18 fragments that bind at similar sites on the NSP14 protein and identifying the overlapping areas of the fragments. Optimization strategies will then be employed to design the new high-molecular-weight molecules. These methods aim to enhance potency and binding affinity, ultimately transforming the compound into a more effective inhibitor.[21]

The images below show clusters of the 18 fragments bound to the NSP14 protein. These clusters exhibit binding at various sites within the NSP14/NSP10 interface. Four distinct clusters were formed due to the overlapping between specific fragments. In Cluster A, overlapping occurs between fragments 2 and 14. Cluster B highlights the overlapping areas between fragments 4, 9, and 17. Cluster C has the largest pocket with significant overlapping, involving fragments 3, 8, 10, 11, 12, 15, and 16. Cluster D illustrates the overlapping between fragments 1 and 6. Notably, fragments 5 and 7 do not exhibit any overlapping with other fragments. Consequently, fragments 5 and 7 will not be considered in the merging of fragments for the design of the new molecule.



4.3 MERGING FRAGMENTS TO DESIGN HIGH-MOLECULAR WEIGHT MOLECULES

In order to design new high-molecular-weight molecules with enhanced binding affinity to a protein, optimization strategies (such as merging, linking, and growing) will be employed in to modify and improve the properties of existing fragments in fragalysis.

Merging involves combining two or more molecular fragments that bind to overlapping regions of the target protein, to create a new larger molecule. This strategy is employed when each fragment contributes specific favourable interactions with the target protein and have structural similarities. By merging these fragments, the resulting molecule may exhibit improved binding affinity and efficacy.

Linking is the process of connecting two or more molecular entities using a linker. The linker connects distinct pharmacophores or functional groups within the molecule. This strategy enables the spatial arrangement of various components to be optimised, resulting in an overall structure with enhanced binding to the target protein.[22]

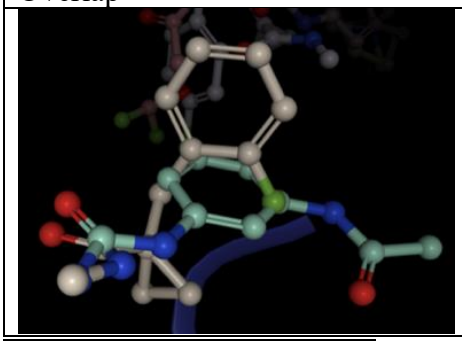
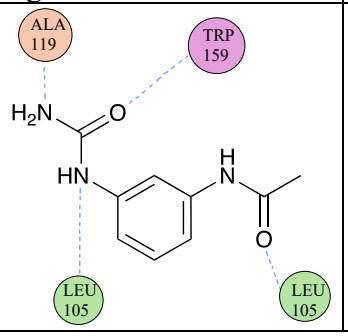
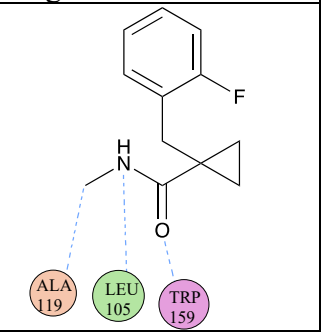
Growing refers to the stepwise addition of functional groups or atoms to an existing molecular scaffold. This incremental expansion aims to explore and exploit novel target-interaction hot spots. Growing is a systematic approach to increasing the size of the molecules with improved binding properties.[22]

In the evaluation of the pharmacokinetic characteristics of the newly designed molecules, the application of both Lipinski's Rule of 5 and the Rule of 3 will be employed. These criteria offer an approach to assess the drug-likeness of compounds based on key physicochemical properties. Its stated that for a fragment to comply with the rule of 3 it will have the following characteristics: a octanol-water partition coefficient log P not greater than 3, a molecular mass less than 300 Daltons, not more than 3 hydrogen bond donors, and not more than 3 hydrogen bond acceptors.[23]

Below I have proposed and designed several molecules by identifying the overlapping fragment pairs and applying the optimising strategies described above.

Creating a molecule/ fragment from the overlapping pairs between:

(1) Fragment 4 and 9

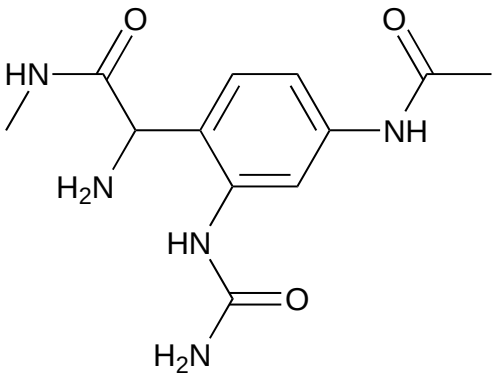
Overlap	Fragment 4	Fragment 9
		

- - - - - -	represents the hydrogen bond
TRP	Tryptophan amino acid
ALA	Alanine amino acid
LEU	Leucine

Observation:

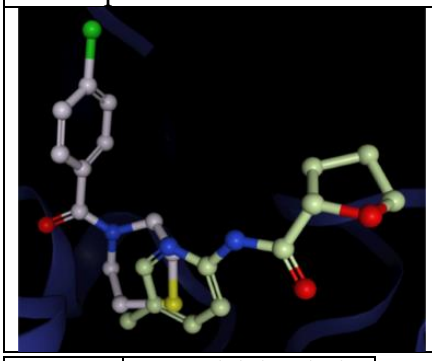
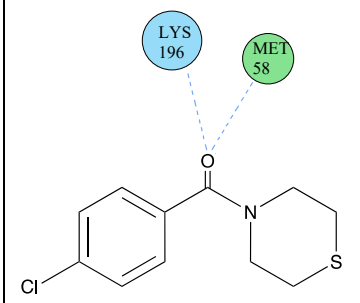
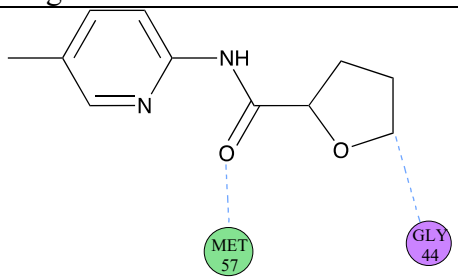
- The C=O group (of fragment 4 and fragment 9) forms a hydrogen bond with the target protein
- The NH_2 group (of fragment 4 and fragment 9) forms a hydrogen bond with the target protein
- The methyl group of fragment 9 forms a weak hydrogen bond with the target protein
- the fluoride group, aromatic ring and cyclohexane group showed no interaction with the protein

suggested fragment:



- This fragment has 7 hydrogen bond donors and 8 hydrogen bond acceptors. Has a molecular weight of 279.30 Da. The LogP is -2.46 (this means that it is hydrophilic/ more soluble in water than in octanol). This follows the rule of 5, but does not follow the 3 of rule as the fragment has more than 3 hydrogen bond donors/ acceptors.
- Retain all the important interactions between the ligand and the target protein, this includes the NH_2 group (as it acts as a H-bond donor) and the C=O group (as the oxygen act as a H-bond acceptor).
- The aromatic ring and fluoride group of fragment 9 has been removed as no interaction with the target protein has been found
- The cyclopropyl group (from fragment 9) does not interact with the target protein. Hence this has been removed and replaced with NH_2 group.

(2) Fragment 15 and 18

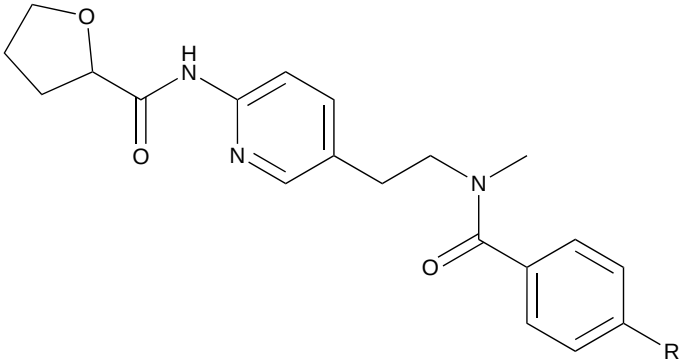
Overlap	Fragment 15	Fragment 18
		

------	represents the hydrogen bond
LYS	Lysine amino acid
MET	Methionine amino acid
GLY	Glycine amino acid

Observation:

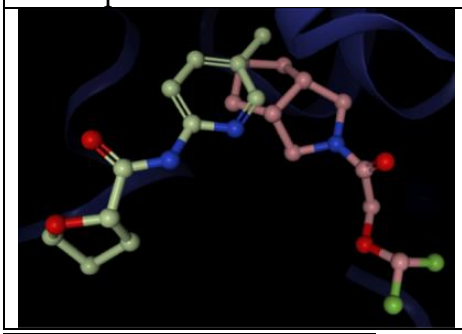
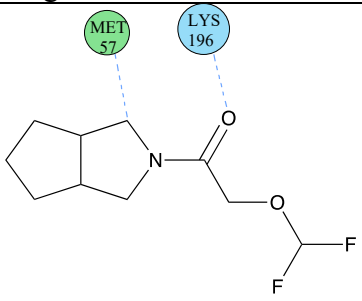
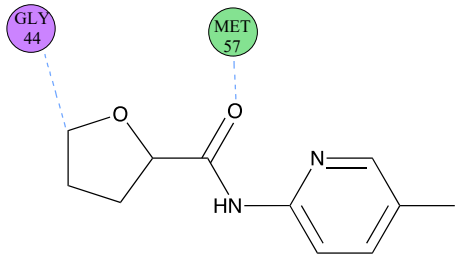
- The amide group (of fragment 15 and 18) forms a hydrogen bond with the target protein
- The methyl of the cyclic ether group (from fragment 18) forms a weak hydrogen bond with the target protein
- There is no interaction between the chloride group (from fragment 15) with the target protein

Suggested molecule design :



- This fragment has 1 hydrogen bond donors and 6 hydrogen bond acceptors. Has a molecular weight of 352.415 Da. The LogP is about 2.33. This follows the Lipinski rule of 5.
- The chloride group and heterocyclic compound (containing sulfur and nitrogen) from Fragment 15 is removed as no interaction is found with the target protein.

(3) Fragment 12 and 18

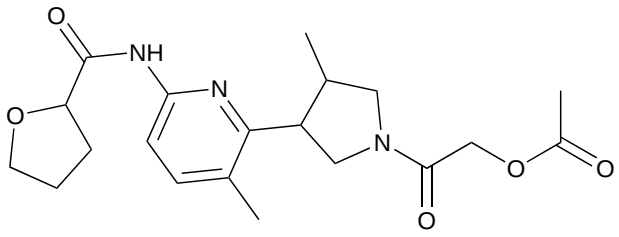
Overlap	Fragment 12	Fragment 18
		

	represents the hydrogen bond
LYS	Lysine amino acid
MET	Methionine amino acid
GLY	Glycine amino acid
AlA	Alanine amino acid

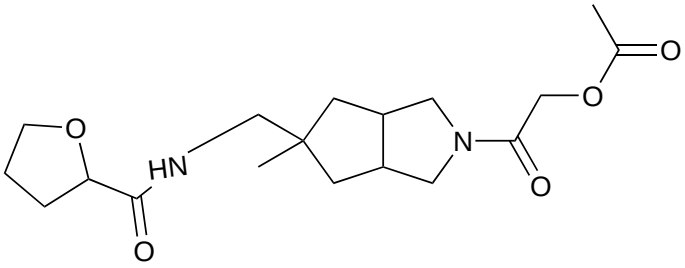
Observation:

- Fragment 12 has 2 interactions with the target protein. One is with the oxygen of the amide group and another interaction is on the carbon of the pyrrolidine group.
- Fragment 18 had 2 interactions with the target protein. One is with the oxygen of the amide group and another interaction is with the methyl group of the cyclic ether group.

Suggested molecule designs :



design 1

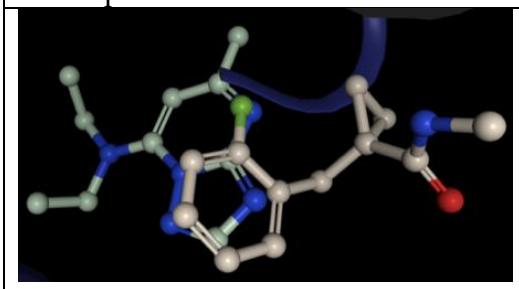
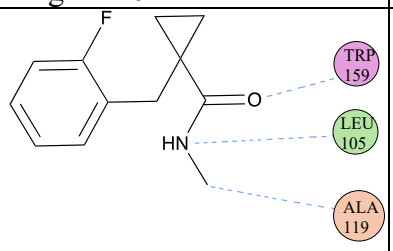
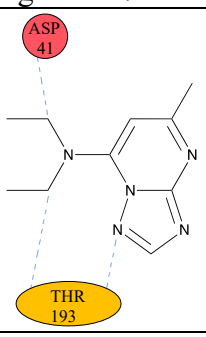


design 2

Design 1 molecule has 1 hydrogen bond donors and 8 hydrogen acceptors. Has a molecular weight of about 389.45 Da. The LogP is about 0.59. This follows the Lipinski rule of 5.
Design 2 molecule has 1 hydrogen bond donors and 7 hydrogen bond acceptors. Has a molecular weight of 352.43 Da. The LogP is about -0.5. This follows the Lipinski rule.

- In both design 1 and 2, The fluoride group (from fragment 12) was removed and replaced with a carbonyl group and a methyl group as it did not interact with the target protein.
- In design 2, the pyridine group was also removed and replaced with a methyl group as it has no interaction with the target protein.

(4) Fragment 9 and Fragment 17

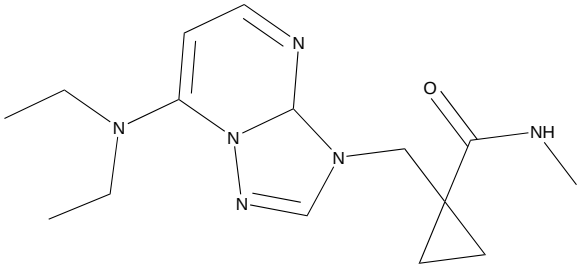
Overlap	Fragment 9	Fragment 17
		

- - - - - -	represents the hydrogen bond
TRP	Lysine amino acid
LEU	Methionine amino acid
ALA	Glycine amino acid
ASP	Aspartic acid
THR	Threonine

Observations:

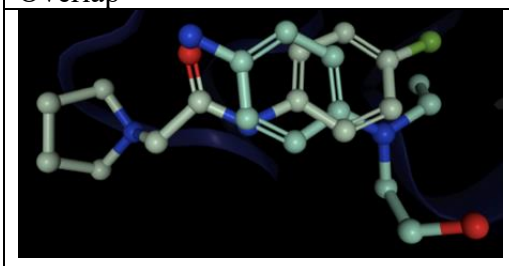
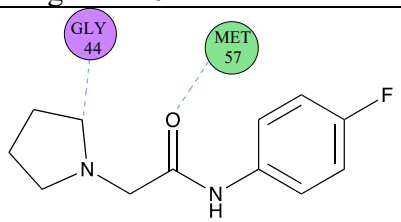
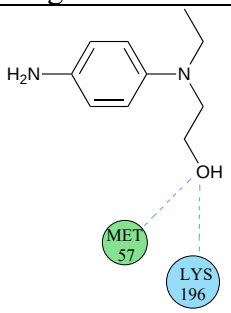
- The methyl of group (from fragment 9) forms a weak hydrogen bond with the target protein. The oxygen and nitrogen from the amide group (of fragment 9) also formed hydrogen bonds with the target protein.
- Fragment 12 has 3 interactions with the protein. One is with the nitrogen of the imidazole group and another is with the methyl group of the tertiary amine

Suggested molecule design :



- This molecule has 1 hydrogen donors and 7 hydrogen acceptors. Has a molecular weight of 304.40 Da. The LogP is about 1.3. This follows the Lipinski rule of 5
- The methyl group (of fragment 17) and the fluoroaromatic ring (of fragment 9) was removed as no interaction was found.

(5) Fragment 10 and Fragment 11

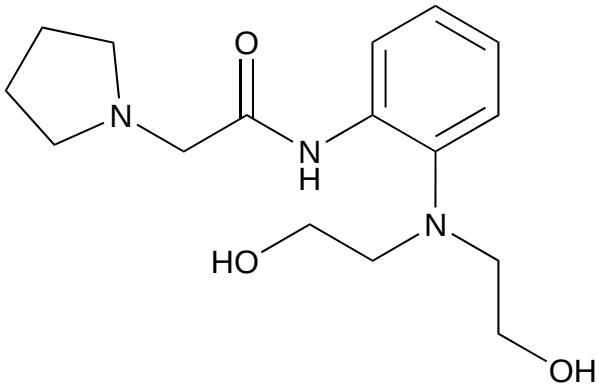
Overlap	Fragment 10	Fragment 11
		

	represents the hydrogen bond
GLY	Glycine
MET	Methionine
LYS	Lysine amino acid

Observation:

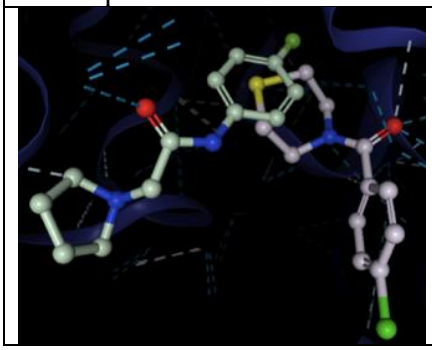
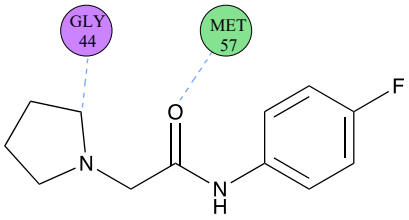
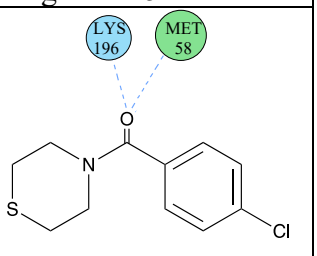
- Fragment 10 has 2 interactions with the target protein. The oxygen of the amide group and the carbon of the pyrrolidine group forms hydrogen bonds with the target protein.
- Fragment 11 has an interaction with the target protein. The oxygen of the hydroxyl (OH) group forms 2 hydrogen bonds with the target protein.
- The fluoride group of Fragment 10 and aromatic amine group of Fragment 11 has no interaction with the target protein

Suggested molecule:



- This fragment has 3 hydrogen bond donors and 6 hydrogen bond acceptors. Has a molecular weight of 307.39 Da. The LogP is about 0.32. This follows the Lipinski rule of 5
- Fluoride group from fragment 10 was removed as it does not interact with the target protein
- Hydroxyl group (OH) was added at the end of the carbon chain (of fragment 11) to replicate the CH_2-CH_2-OH chain (of fragment 11) as it interacts with the target protein.

(6) Fragment 10 and 15

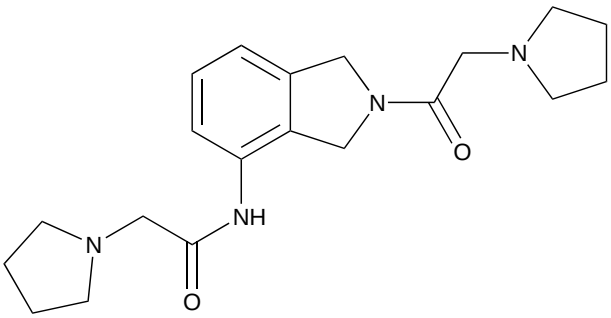
Overlap	Fragment 10	Fragment 15
		

- - - - - -	represents the hydrogen bond
GLY	Glycine
MET	Methionine
LYS	Lysine

Observation:

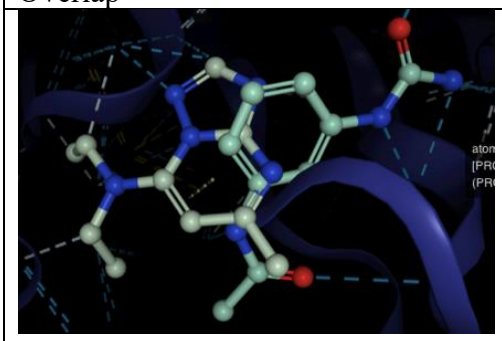
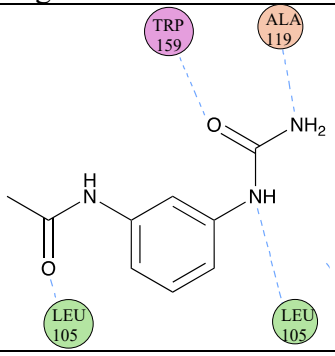
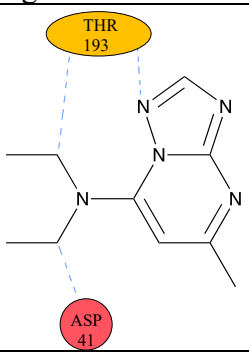
- Fragment 10 has 2 interactions with the target protein. One with oxygen of the amide group and another with the methyl group of the pyrrolidine group
- The oxygen of the amide group (from fragment 15) forms hydrogen bonds with target protein
- The fluoride group, sulphide group and chloride group has no interaction with the protein, hence these functional groups are not included in the design of the suggested molecule

Suggested molecule design:



- This molecule has 1 hydrogen bond donors and 6 hydrogen bond acceptors. Has a molecular weight of 356.47 Da. The LogP is about 0.12. This follows the Lipinski rule of 5.
- The S-N group and aryl chloride from fragment 15 were both removed and replaced with a pyrrolidine group to replicate fragment 10.
- The fluoride group from fragment 10 was removed as no interaction was found.

(7) Fragment 17 and Fragment 4

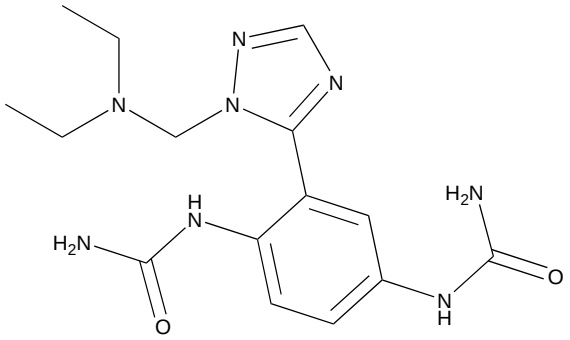
Overlap	Fragment 4	Fragment 17
		

- - - - - -	represents the hydrogen bond
TRP	Tryptophan
LEU	Leucine
THR	Threonine
ASP	Aspartic acid
ALA	Alanine

Observation:

- Fragment 4 has 4 interactions with the protein. 2 interaction with the oxygen of the amide group and 2 interactions with the nitrogen of the amide group
- The methyl of the amine group (from fragment 17) forms a weak hydrogen bonds with the target protein. The nitrogen of the imidazole group also formed hydrogen bond with the target protein.
- There's no interaction with the aromatic ring or pyrimidine group

suggested molecule design :



- This molecule has 6 hydrogen bond donors and 10 hydrogen bond acceptors. Has a molecular weight of 346.40 Da. The LogP is about -0.01 (this means that the suggested molecule is more soluble in water than octanol, it is hydrophilic). This follows the Lipinski rule of 5
- The methyl group of fragment 4 was removed and replaced with an amine group.
- The aromatic ring containing nitrogen from fragment 17 was removed as no interaction was found.

5. Discussion

By conserving interactions through merging of fragments, redundant chemical entities are eliminated, reducing complexity and enhancing the overall simplicity of the molecular design.

The analysis of the 18 fragments revealed that the halide groups (such as chloride and fluoride) present in the fragments did not participate in ligand-protein interactions, leading to their omission in the design of the proposed molecules. The majority of fragments featuring functional groups like carbonyl, hydroxyl, and amides demonstrated (hydrogen bond) interactions with the target protein. It was also observed that fragments containing methyl groups attached to nitrogen atoms formed weak hydrogen bonds with the target protein.

While molecules can interact with proteins through various mechanisms in ligand-protein interactions, no other intermolecular forces, such as hydrophobic interactions, were observed between the 18 fragments and the amino acid residues in the target protein. The effective and high-affinity binding of a protein to a ligand is contingent upon the simultaneous formation of weak, noncovalent bonds, including hydrogen bonds, ionic bonds, and van der Waals attractions, along with favourable hydrophobic interactions. Given the inherent weakness of each individual bond, successful binding necessitates the simultaneous formation of numerous weak bonds.[24]

Referring to the NSP14/10 complex, the interaction interface involves the formation of a β-sheet, hydrogen bonds, salt bridges, and hydrophobic interactions between NSP14 and NSP10.[19]

Therefore, comprehending the diverse array of interactions in ligand-protein binding is crucial for advancing our understanding of how molecules can tightly bind to the target protein.

In this study, we have designed molecules (by merging the overlapping pairs of fragments) with the aim of their potential to serve as therapeutic drugs. However, synthesizability is a crucial factor that requires careful consideration in molecular design. This means that we may need to modify the structures to facilitate easier synthesis. Our goal is to promptly obtain suggestions using commercially available building block. This approach aims to avoid spending months on synthesizing a molecule that may prove ineffective or fail to bind to the target protein as intended.

6. Conclusion

In conclusion, we have analysed 18 smaller molecules known as fragments, identified the overlapping pairs of fragments and employed optimization strategies (such as merging, linking, and growing) to design seven new molecules. These methods aim to enhance potency and binding affinity, ultimately transforming the compound into a more effective inhibitor.[21]

In summary, further research and investigation are still needed to assess the feasibility of synthesising and measuring the effectiveness of these newly designed molecules as potential therapeutic drug candidates. However, this ongoing exploration has significantly expanded the possibilities and opportunities for molecular design, paving the way for continued advancements in the field.

7. Future work

In the future, our plan involves virtually testing the newly designed molecules, synthesizing the most feasible ones, and then evaluating their efficacy. This evaluation will be conducted by integrating insights from computational modelling, chemical synthesis, and experimental validation.

Experimental validation such as biological activity testing in enzyme assays and cell-based antiviral assays, should be conducted with all the selected molecules to identify a tighter binder and understand the efficacy, safety, and mechanism of action of the molecules.

We want to be able obtain physical data to prove their function and demonstrate whether the newly designed molecule can effectively inhibit the activity of NSP10 and NSP14, preventing viral replication.

Another consideration is that Protein-Protein Interaction (PPI) binding sites often lack well-defined pockets, posing a challenge in identifying small molecules that can strongly and specifically target them.[18] Therefore, it is necessary to assess the size or capacity of the specific binding sites on the NSP14 proteins where the fragments are interacting. This involves analysing the space or pocket on the protein's surface where the fragments bind and evaluating whether this volume is sufficient to accommodate a small molecule capable of tight binding.

Our entire body of work has been uploaded online through a platform called GitHub, serving as an open invitation to the scientific community to actively engage in the field of molecular design. Seeking collaboration with external experts or research groups with specific knowledge in the therapeutic area will provide us with valuable insights, diverse perspectives, and complementary expertise. This collaborative approach is envisioned to facilitate the seamless continuation of our project by future contributors.

REFERENCES:

1. Hu B., Guo H., Zhou P., Shi Z L., Characteristics of SARS-CoV-2 and COVID-19, *Nat Rev Microbiol*, **2021**, *19*, 141–154 DOI: https://doi.org/10.1038/s41579-020-00459-7
2. Yue W M., Rong Z., Juan G L., Fei G X., Ping W D., Min C J., SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development, *Frontiers in Cellular and Infection Microbiology*, **2020**, *10*, 1-14 https://doi.org/10.3389/fcimb.2020.587269
3. Štefek M., Chalupská D., Chalupský K., Zgarbová M., Dvořáková A., Krafčíková P., Li A S M., Šála M., Dejmek M., Otava T., Chaloupecká E., Kozák J., Kozic J., Vedadi M., Weber J., Kaiserová H M., Nencka R., Rational Design of Highly Potent SARS-CoV-2 nsp14 Methyltransferase Inhibitors, *ACS Omega*, **2023**, *8*, 27410-27418 DOI: 10.1021/acsomega.3c02815
4. Walls A C., Park Y J., Tortorici M A., Walls A., McGuire A T., Veesler D., Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein, *cell press*, **2020**, *2*, 281–292 DOI: 10.1016/j.cell.2020.02.058
5. Hassan H., Chiavaralli J., Hassan A., Bedda L., Krischuns T., Chen Y K., Li A S M., Delpal A., Decroly E., Vedadi M., Naffakh N., Agou F., Mallart S., Arafa R K, Arimondo P B., Design and Synthesis of naturally-inspired SARS-CoV-2 Inhibitors, *RSC Med Chem*, **2023**, *14*, 507-519 https://doi.org/10.1039/D2MD00149G
6. Canal B., McClure A W., Curran J F, Wu M., Ulferts R., Weissmann F., Zeng J., Bertolin A P., Milligan J C., Basu S., Drury L S., Deegan T D., Fujisawa R., Roberts E L., Basier C., Labib K., Beale R., Howell R., Diffley J F X., Identifying SARS-CoV-2 Antiviral Compounds by Screening for Small Molecule Inhibitors of nsp14/nsp10 Exoribonuclease, *Biochem J*, **2021**, *13*, 2445–2464. DOI: 10.1042/BCJ20210198
7. Ma Y., Wu L., Shaw N., Gao Y., Wang J., Sun Y., Lou Z., Yan L., Zhang R., Rao Z., Structural Basis and Functional Analysis of the SARS Coronavirus nsp14-nsp10 Complex, *Proceedings of the National Academy of Sciences of the United States of America*, **2015**, *30*, 9436–9441. DOI: 10.1073/pnas.1508686112
8. Kozielski F., Sele C., Talibov V O., Lou J., Dong D., Wang Q., Shi X., Nyblom M., Rogstam A., Krojer T., Fisher Z., Knecht W., Identification of Fragments Binding to SARS-CoV-2 nsp10 reveals ligand-binding sites in conserved interfaces between nsp10 and nsp14/nsp16, *RSC Chemical Biology*, **2022**, *3*, 44-55 https://doi.org/10.1039/D1CB00135C
9. Saramago M., Bárria C., Costa V G., Souza C S, Viegas S C., Domingues S., Lousa D., Soares C M., Arraiano C M., Matos R G., New targets for drug design: importance of nsp14/nsp10 complex formation for the 3'-5' exoribonucleolytic activity on SARS-CoV-2, *The Febs Journal*, **2021**, *17*, 5130-5147 doi: 10.1111/febs.15815
10. Decroly E., Imbert I., Coutard B., Bouvet M., Selisko B., Alvarez K., Gorbalenya A E., Snijder E J., Canard B., Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity, *J Virol*, **2008**, *82*, 8071− 8084 DOI: 10.1128/JVI.00407-08
11. Bouvet M., Lugari A., Posthuma C C., Zevenhoven J C., Bernard S., Betzi S., Imbert I., Canard B., Guillemot J C., Lécine P., Pfefferle S., Drosten C., Snijder E J., Decroly E., Morelli X., Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes, *The Journal of biological chemistry*, **2014**, *37*, 25783–25796 doi: 10.1074/jbc.M114.577353
12. Chen J., Zhou Y., Wei X., Xu X., Qin Z., Ong C P., Ye Z W., Jin D Y., Boitrel B., Yuan S., Chan J F W., Li H., Sun H., Development of Pan-Anti-SARS-CoV-2 Agents through Allosteric Inhibition of nsp14/nsp10 Complex, *ACS Infectious Diseases Article ASAP*, **2023**, DOI: 10.1021/acsinfecdis.3c00356
13. Minter B M., Vries M D., Kaur R., Schinlever A., Garcia S F., Goldberg H V., Wang H., Hinds T R., Bailly F., Zheng N, Cotelle P., Desmaële D., Landau N R., Dittmann M., Pagano M., The NSP14/NSP10 RNA repair complex as a Pan-coronavirus therapeutic target, *Cell Death Differ*, **2022**, *29*, 285–292 doi: 10.1038/s41418-021-00900-1
14. Baddock H T., Brolih S., Yosaatmadja Y., Ratnaweera M., Bielinski M., Swift L P., Migoni A C., Fan H., Keown J R, Walker A P., Morris G M., Grimes J M, Fodor E., Schofield C J., Gileadi O., McHugh P J., Characterization of the SARS-CoV-2 ExoN (nsp14ExoN–nsp10) complex: implications for its role in viral genome stability and inhibitor identification, *Nucleic Acids Research*, **2022**, *3*, 1484–1500 https://doi.org/10.1093/nar/gkab1303
15. Basu S., Mak T., Ulferts R., Wu M., Deegan T., Fujisawa R., Tan K W., Lim C T., Basier C., Canal B., Curran J F., Drury L S., McClure A W., Roberts E L., Weissmann F., Zeisner T U., Beale R., Cowling V H., Howell M., Labib K., Diffley J F X., Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of Nsp14 RNA cap methyltransferase, *Biochem J*, **2021**, *13*, 2481–2497 doi: 10.1042/BCJ20210219
16. Otava T., Šála M., Li F., Fanfrlík J., Devkota K., Perveen S., Chau I., Pakarian P., Hobza P., Vedadi M ., Boura E., Nencka R., The Structure-Based Design of SARS-CoV-2 nsp14 Methyltransferase Ligands Yields Nanomolar Inhibitors, *ACS Infectious Diseases*, **2021**, *7*, 2214-2220 doi: 10.1021/acsinfecdis.1c00131
17. Scott D., Bayly A., Abell C. Small molecules, big targets: drug discovery faces the protein–protein interaction challenge. *Nat Rev Drug Discov,* **2016**, *15*, 533–550. https://doi.org/10.1038/nrd.2016.29
18. Thurston D E., Zinzalla G., Targeting protein–protein interactions for therapeutic intervention: a challenge for the future, Future Medicinal Chemistry, **2009**, *1*, 65-93 https://doi.org/10.4155/fmc.09.12

This project is a Type C Data-Analysis with a word count of 5624 words formatted in the printed style of Journal of Medicinal Chemistry

19. Imprachim N., Yosaatmadja Y., Newman J A., Crystal structures and fragment screening of SARS-CoV-2 NSP14 reveal details of exoribonuclease activation and mRNA capping and provide starting points for antiviral drug development, *Nucleic Acids Research*, **2023**, *1*, 475–487 DOI: https://doi.org/10.1093/nar/gkac1207
20. Skyner R., What is Fragalysis? (An introduction), *Diamond*, **2020**, https://diamondlightsource.atlassian.net/wiki/spaces/FRAG/pages/8356087/What+is+Fragalysis+An+introduction
21. Oleinikovas V., Saladino G., Cossins B P., Gervasio F L., Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations, *Journal of the American Chemical Society*, **2016**, *43*, 14257-14263. DOI: 10.1021/jacs.6b05425
22. Kirsch P., Hartman AM., Hirsch AKH., Empting M., Concepts and Core Principles of Fragment-Based Drug Design, *Molecules*, **2019**, *24*, 4309. https://doi.org/10.3390/molecules24234309
23. Goodwin R J A., Bunch J., McGinnity D F., Mass Spectrometry Imaging in Oncology Drug Discovery, *Academic Press*, **2017**, *134*, 133-171, DOI: https://doi.org/10.1016/bs.acr.2016.11.005
24. Alberts B., Johnson A., Lewis J., Raff M., Roberts K., Walter P., Protein Function, Garland Science, 2002, 4. Available from: https://www.ncbi.nlm.nih.gov/books/NBK26911/