



Open Source Drug Discovery: Fragment-Based Drug Discovery of NSP13 Inhibitors in Treatment of COVID-19

Kangping Liu

A dissertation submitted in partial fulfilment of the requirements
for the degree of

Master of Science

of

University College London

Supervised by:

Professor Matthew H. Todd

School of Pharmacy, UCL

August 2022

Table of Contents

- i. Acknowledgements
- ii. Plagiarism Statement
- iii. Abstract

1 Introduction

- 1.1 COVID-19 Pandemic and SARS-CoV-2
 - 1.1.1 Impact of SARS-CoV-2 and Its Structure
 - 1.1.2 Lifecycle of SARS-CoV-2
 - 1.1.3 Pathophysiology of SARS-CoV-2 Infection
- 1.2 Treatment for COVID-19
 - 1.2.1 FDA Approved Drugs
 - 1.2.2 Discovery of New Drugs
- 1.3 SARS-CoV-2 viral RNA Replication Machinery
 - 1.3.1 Non-Structural Proteins (NSPs) and Replication Transcription Complex
 - 1.3.2 NSP13
 - 1.3.3 NSP13 Structure and Function Relationship
- 1.4 Small Molecule Binding of Non-Structural Protein 13
 - 1.4.1 Binding Pocket Search
- 1.5 Aim of Project

2 Results and Discussion

- 2.1 Structural Based Drug Design of NSP13 Inhibitors
 - 2.1.1 Binding Analysis in C-Terminus-B
 - 2.1.2 Autodock Vina Docking results
- 2.2 Synthesis of Fragments in Initial Screening
 - 2.2.1 Synthesis of 4-Fluoro-N-[(4-fluorophenyl)methyl]benzenacetamide
 - 2.2.2 Synthesis of 5-Fluoro-3-(2-fluorophenyl)-1*H*-indene-1,2(3*H*)-dione

2.2.3 Synthesis of 3-(4-Fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile

3 Conclusion and Future Work

4 Experimental

5 Bibliography

6 Appendix

6.1 Supplementary information

Acknowledgments

First, I would like to show my gratitude to Prof. Matthew Todd for the opportunity to work in the Todd group in school of pharmacy. His help in both reactions and thesis writing enlightened and inspired me.

Second, I would like to express my thanks for other members in Todd group. They have taught me a lot in using different instruments including LCMS, NMR spectroscopy and biotage etc. They are so warm-hearted and more than willing to help.

Additionally, I would like to show my thanks to other members in G25. They helped me a lot in looking for reagents.

Last but not least, I would like to thank my parents for their support in all aspects, especially in my education abroad.

Plagiarism Statement

I confirm that the work presented in this dissertation is my own. Where information has been derived from other sources, I confirm that this has been indicated in the dissertation.

Abstract

Coronavirus disease 2019 (COVID-19) has impacted the life of every human. SARS-CoV-2 has been found to be the pathogen of COVID-19. SARS-CoV-2 has 16 non-structural proteins (NSPs) that are involved in its lifecycle and these NSPs have different functions. Among 16 NSPs, NSP 13 is a helicase and is the most conserved NSP in coronavirus. NSP13 is involved in the viral RNA replication, proof-reading and backtracking mechanisms, so successful inhibition of NSP13 can interrupt the viral replication and ameliorate the SARS-CoV-2 infections. By far, no drugs targeting NSP13 has been approved so discovery of NSP13 inhibitors is necessary for not only SARS-CoV-2 infections but also other coronavirus treatments. The conformational change of NSP13 indicate the inhibition of NSP13 can be potentially achieved by targeting an allosteric site and this site has been found, which was named as C-terminus-B. In addition, 5 fragments that can bind to C-terminus-B have been reported in fragment-bound NSP13 crystal structures. A pharmacophore was summarised from crystal structure and binding analysis of those 5 fragments and this pharmacophore was used for fragment screening. The screening generated a series of fragments, and the fragments are ranked according to their binding scores. 3 of 30 top-score fragments were targeted to synthesis in this project.

1 Introduction

1.1 COVID-19 Pandemic and SARS-CoV-2

1.1.1 Impact of SARS-CoV-2 and Its Structure

Coronavirus disease 2019 (COVID-19), a respiratory disease first reported at the end of 2019 in Wuhan, China, has impacted the world and caused more than 6.4 million deaths globally by 5 Aug 2022. (Zhu et al., 2020, WHO, 2022) The pathogen of COVID-19 is severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a beta-coronavirus that is similar to those found in bats. (Zhu et al., 2020) Infection of SARS-CoV-2 can cause symptoms ranging from mild respiratory symptoms to severe organ failure or even multiple organ failure. (Yuki et al., 2020)

SARS-CoV-2 is an enveloped RNA virus, which is icosahedral symmetric, with a single-stranded positive RNA genome. (Weiss and Navas-Martin, 2005) The virial genome RNA consists of 14 open reading frames (ORFs), and 2 of them, ORF1a and 1b, take up approximately 67 % of the genome and encode for 2 polyproteins, 1a and 1b (pp1a and pp1b); while the other ORFs are relevant to synthesis of spike, membrane, nucleocapsid and other accessory proteins. (Helmy et al., 2020, Alanagreh et al., 2020) Spike proteins can recognise receptors on host cells. (Demogines et al., 2012) Membrane proteins will shape the membranes and support virions. (Neuman et al., 2011) Envelope proteins are involved in the assembly and release of virions.(Alanagreh et al., 2020, DeDiego et al., 2007) Nucleocapsid proteins are relevant to RNA genome packaging and pathogenicity when acting as an interferon inhibitor. (Cui et al., 2015, Alanagreh et al., 2020) The polyprotein 1a and 1b will be further processed into 16 non-structural proteins (NSPs). (Chen et al., 2020c) The replication-transcription complex (RTC), which is responsible for virial genome replication and transcription, is formed by NSPs. (Chen et al., 2020c)

1.1.2 Lifecycle of SARS-CoV-2

In a SARS-CoV-2 infection (Figure 1), the virus first binds to the receptor, angiotensin-converting enzyme 2 (ACE2) of lung cells using its spike proteins. (Hoffmann et al., 2020, Li et al., 2003, Qian et al., 2013, Alanagreh et al., 2020) After binding to surface receptors, cell surface-associated transmembrane protease serine 2 (TMPRSS2) and cathepsin will then cleave the trimer spike protein. (Hoffmann et al., 2020) Subunit S1 of the spike protein contributes to attachment to the host and S2 is responsible for the membrane fusion and endocytosis of the virus. (Hoffmann et al., 2020) After the virus entry, the viral RNA is then released to cytoplasm, and pp1a and pp1b are then translated and processed into NSP 1-16 and ultimately assemble to the RTC. This complex can help with production of negative sense RNA from template genomic RNA and then more genomic RNA and subgenomic RNA (sgRNA), which is for translation of structural proteins for virion assembly. sgRNA can have several ORFs, but only the one closest to the 5' end will be translated. Viral structure proteins except nucleocapsid proteins are translated at host endoplasmic reticulum (ER), so nucleocapsid proteins will enter the ER-Golgi intermediate compartment after their production in cytoplasm. Ultimately, the mature virion will bud off from the host cell membrane. (Alanagreh et al., 2020)

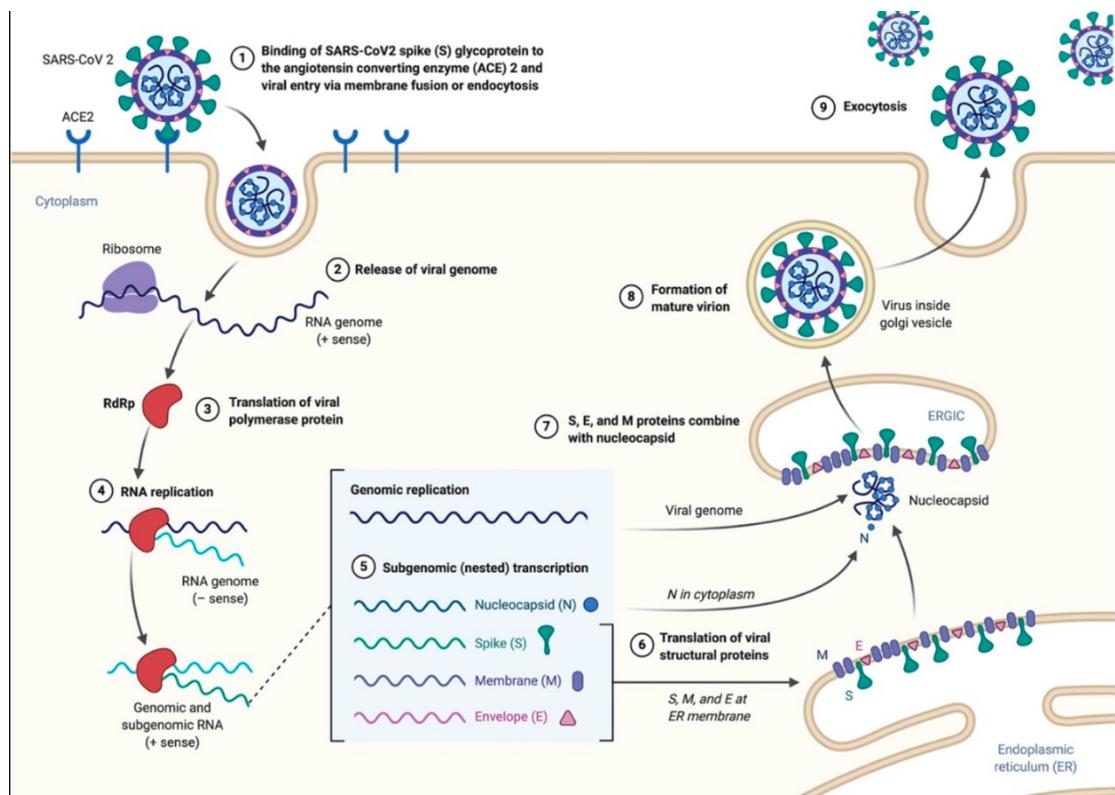


Figure 1. The lifecycle of SARS-CoV.(Alanagreh et al., 2020) (Use of figure allowed under Creative Commons Attribution License)

1.1.3 Pathophysiology of SARS-CoV-2 Infection

The lung epithelial cells are primarily susceptible to SARS-CoV-2 because large quantities of ACE2 are expressed on lung epithelial cells, so the entry of virus is more likely by binding these receptors. (Jia et al., 2005, Hamming et al., 2004) After infection, some severe COVID-19 patients show immunological symptoms such as lymphopenia, especially for T cells in the peripheral blood. (Zhou et al., 2020, Qin et al., 2020) This reduction of T cell is likely to facilitate the progress of the disease. (Yuki et al., 2020) Studies have shown the correlation between the expression level of NKG2A receptors and the count of natural killer (NK) cell and CD8⁺ T cells. A high NKG2A expression level and a low count of those cytotoxic lymphocytes are observed in COVID-19 patients. (Zheng et al., 2020) CD4⁺ and CD8⁺ T cells of severe patients are also found activated and showed a high level of proinflammatory cytokines including

interleukin-6 (IL-6) in plasma. (Huang et al., 2020, Qin et al., 2020, Zhou et al., 2020, Yuki et al., 2020) Activated CD4⁺ T cells will become T helper (Th) 1 cells and produce granulocyte-macrophage colony-stimulating factor (GM-CSF) and IL-6 is highly expressed by cytokine induced CD14⁺ CD16⁺ monocytes, which will accelerate the inflammation. (Zhou et al., 2020) Pathological studies have confirmed that large quantities of inflammatory-cell infiltrations happened in the lung of severe patients. (Xu et al., 2020, Tian et al., 2020) Therefore, the Th 1 cells are likely to enter the pulmonary circulation and are detrimental to lung functions and lead to death of severe patients. (Zhou et al., 2020) In addition, high level of d-dimer and fibrinogen is also observed in severe patients, and this can account for non-respiratory symptoms such as thrombosis and pulmonary embolism after infection of SARS-CoV-2. (Yuki et al., 2020) Since endothelium can facilitate vasodilation and fibrinolysis, reduce aggregation, and regulate thrombosis, the thrombophilia of severe patients may indicate damage to endothelium. (Wang et al., 2018a, Yuki et al., 2020) However, nearly 33% of the lung is comprised of endothelial cells, so endothelial injury can increase the microvascular permeability and ultimately the invasion of SARS-CoV-2. (Zeng et al., 2012, Yuki et al., 2020)

1.2 Treatment for COVID-19

Currently, FDA has approved 4 drugs for COVID-19, remdesivir, baricitinib, Paxlovid and molnupiravir. (FDA, 2022) Remdesivir is a phosphoramidate prodrug in a form of an adenosine monophosphate (AMP) analog and will be hydrolysed to monophosphate GS-441524 and then phosphorylated to its active form of the corresponding triphosphate (GS-443902) *in vivo*. (Jorgensen et al., 2020a, Singh et al., 2020) The masked phosphate in the prodrug is presumed to enable the drug to enter cells faster than corresponding compound containing charged phosphates. (Jorgensen et al., 2020a) GS-443902 is a competitive inhibitor of endogenous ATP and can thus inhibit the replication of

SARS-CoV-2. (Jorgensen et al., 2020a) When GS-441524 is incorporated in the viral RNA, it can lead to an early termination of replication. (Douedi and Miskoff, 2020) The structural difference between GS-441524 and adenosine triphosphate (ATP) is an extra cyano group at the 1' position of GS-441524. This cyano group makes both GS-441524 and GS-443902 more selective to RNA dependent RNA polymerase (RdRp), the viral RNA polymerase, than the polymerases in humans. (Jorgensen et al., 2020a)

A particular challenge in developing nucleoside mimicking drugs for coronaviruses is the proof-reading mechanism of exoribonuclease (ExoN) because it can correct the errors in RNA synthesis. (Jorgensen et al., 2020a, Agostini et al., 2018) However, Remdesivir can evade the proof-reading mechanism of ExoN to some extent, and this can be achieved by preference for GS-442514 of RdRp and delayed chain termination of RNA. (Jorgensen et al., 2020a) Incorporation of GS-442514 in RNA can lead to a delayed chain termination because it has a free 3'-hydroxyl group and endogenous nucleosides can add on it. (Jorgensen et al., 2020a) Therefore, the RNA structure is disrupted, and elongation of RNA is stopped. (Shannon et al., 2020) After incorporating of GS-442514 in RNA chain, 3 more nucleotides will be added, and it is supposed that these 3 nucleotides help with the evasion of proof-reading and error correction mechanisms. (Gordon et al., 2020a, Gordon et al., 2020b) Mutation of F476L and V553L in RdRp in beta-coronavirus murine hepatitis virus cell culture confers 2.4-fold and 5-fold resistance respectively in vitro and 5.6-fold if both are mutated. (Agostini et al., 2018) However, the EC₅₀ values can still be achieved within human drug exposure safe limit. (Jorgensen et al., 2020a) In addition, the wild-type virus can outcompete against the mutant without the effects of remdesivir. (Jorgensen et al., 2020a) Further studies specific to SARS-CoV-2 are still needed to illustrate the drug resistance.

Baricitinib is an inhibitor of Janus Kinase (JAK). (Alanagreh et al., 2020) JAK/signal transducers and activators of transcription (STAT) pathway is involved in immune response regulations. (Fragoulis et al., 2019, Ghoreschi et al., 2009) JAK/STAT pathway is triggered by the binding of cytokines to their corresponding receptors on the cell surface. (Liu et al., 1998) This binding leads to conformational change of JAK complexes and their activation. (Liu et al., 1998) The activation of JAK complexes is followed by their autophosphorylation, which makes them more active, and phosphorylation of the cytokine receptors, which can help with the docking of STAT molecules to the receptors. (Liu et al., 1998) After docking of STAT molecules, JAKs will phosphorylate them, and the phosphorylated STAT molecules will leave the receptors, form dimers, travel to nucleus and trigger transcription of certain genes. (Liu et al., 1998) After inhibition of JAKs, when IL-6 or GM-CSF, either of which is highly produced in SARS-CoV-2 severe patients as mentioned in section 1.1, binds to their corresponding receptors, no signal transduction will be triggered. (Liu et al., 1998) Signal transduction of IL-6 is associated with JAK1 and 2. (Jorgensen et al., 2020b) Baricitinib preferentially binds to JAK1 and 2 and its potency for JAK3 is lower, which can still maintain the immune response at a certain level. (Jorgensen et al., 2020b) Interestingly, another pathway of baricitinib inhibition is also proposed in treatment of the cytokine storm in SARS-CoV-2 infections. Baricitinib may inhibit AP2-associated protein kinase 1 (AAK1) so can disrupt the endocytosis and then viral cell entry and intracellular formation of virion are reduced. (Richardson et al., 2020) However, this AP2-AAK1 pathway is predicted by BenevolentAI which integrates known medical informations so validity needs more data to prove.

Different from intravenous drugs remdesivir and baricitinib, Paxlovid and molnupiravir are administered orally. (FDA, 2022) Paxlovid is a combination of 2 drugs and is composed of nirmatrelvir and ritonavir. (Drozdal et al., 2021)

Nirmatrelvir is an inhibitor for 3CL protease. Nirmatrelvir is a tri-peptide and has a moiety that is mimicking glutamine because 3CL protease will preferentially cleave at a glutamine. (Marzi et al., 2022) Nirmatrelvir can form a covalent bond with 3CL protease and then block the production of NSPs in the virus. Ritonavir can inhibit the cytochrome P450 (CYP450) 3A and CYP2D6. (Drozdal et al., 2021) CYP450 3A and CYP2D6 are metabolic enzymes and inhibition of activity will cause a reduction in metabolism. Therefore, Nirmatrelvir will be less likely metabolised and effective for longer time. Studies have shown that Paxlovid can lower the risks of hospital entry and death by 89%. (Marzi et al., 2022) Molnupiravir is a pro-drug that can be hydrolysed to its active form of a nucleoside analog in vivo. (Pourkarim et al., 2022) The active form of molnupiravir is preferred by RdRp over uridine or cytidine-triphosphate and thus RNA replication is disrupted. The infection can be ameliorated by disruption of viral RNA synthesis. None of FDA approved drugs target NSP13 by far, so NSP13 inhibitors are in demand in treatment of SARS-CoV-2 infections.

1.2.1 Discovery of New Drugs

In addition to the drugs approved for treatment of SARS-CoV-2 infections, there is still space for other types of drugs or drugs focusing on other targets. For example, monoclonal antibodies specific for IL-6 or GM-CSF can reduce the cytokine storm and inflammation, and ameliorate disease symptoms. (Zhou et al., 2020) Drugs targeting NKG2A can ameliorate the exhaustion of cytotoxic lymphocyte functions and thus symptoms of COVID-19 (Zheng et al., 2020) In principle, selective inhibition of viral targets is more desirable because such a mechanism is less likely to interfere with normal human biological functions, allowing inhibition of viral replication with minimal off-target effects. New inhibitors targeting other parts of SARS-CoV-2 replication machinery, especially NSP13, are therefore still in demand.

1.3 SARS-CoV-2 Viral RNA Replication Machinery

1.3.1 Non-Structural Proteins and Replication Transcription Complex

The SARS-CoV-2 viral RNA replication machinery is composed of a series of functional and regulating NSPs. (Romano et al., 2020) Functional NSPs involve NSP12, known as RdRp; NSP13, a helicase; NSP14, responsible for mRNA capping and proof-reading; NSP15, a uridylate-specific endoribonuclease and counteracting sensing of double strand RNA; and NSP16, a 2'-O-ribose methyl transferase related to mRNA capping. (Romano et al., 2020) These NSPs are the major enzymes in the replication and transcription of the viral genome. (Arya et al., 2021) Regulating NSPs consist of NSP7 and 8, cofactors of NSP12; NSP9, a protein binding to RNA; and NSP10, a cofactor of NSP14 and 16. (Romano et al., 2020) NSP1 has been shown to have the ability of binding to the 40S ribosomal subunit and cleaving mRNA of the host cell, and thus disrupting the protein translation of the host cell. (Kamitani et al., 2009) This disruption can reduce the host innate immune response because the expression of host factors, which are supposed to defend against the infection of SARS-CoV-2, is undermined. (Arya et al., 2021) NSP3 is a protein anchored to the membrane and interact with other proteins and contributes to the assembly of RTC. (Angelini et al., 2013, Imbert et al., 2008, Arya et al., 2021) NSP5 acts as a cysteine protease, and it can process the viral polyprotein into 12 functional proteins. (Lee et al., 1991, Ziebuhr et al., 2000) The knowledge of functions of other NSPs still remains unclear and further studies are needed. The RTC is composed of NSP2-16 and other factors, and it is necessary in production of viral RNA. (Arya et al., 2021, Romano et al., 2020)

1.3.2 NSP13

Studies have reported that NSP13 plays a necessary role in viral RNA

replication. (Jia et al., 2019, Kleymann et al., 2002, Seybert et al., 2005) In addition, structural mapping of NSP13 showed that the similarity between NSP13 and human proteins IGHMBP2 and UPF1 suggests a potential role of NSP13 of reducing host immunological functions. (O'Donoghue et al., 2021) Therefore, the inhibition of NSP13 can potentially ameliorate the symptoms of SARS-CoV-2 patients in different aspects (viral function and human response) concurrently. NSP13 is thought to be a promising target for SARS-CoV-2 drugs, and there has been a significant amount of research on the druggability of NSP13. (Cavasotto et al., 2021, Habtemariam et al., 2020, Arya et al., 2021) NSP13 of SARS-CoV-2 is highly similar to that of SARS-CoV, and the only difference is the residue V570I. (Newman et al., 2021, Arya et al., 2021) Therefore, NSP13 is presumably conserved in other coronaviruses. (Newman et al., 2021) This conservation of sequence and thus structure of NSP13 makes its inhibitor a potentially wide spectrum drug against the virus family. The NTPase ability, capping ability on the 5' end of RNA, and DNA and RNA duplex unwinding activity of NSP13 have been found. (Shum and Tanner, 2008, Neuman et al., 2014, Chen et al., 2020b) NSP13 is a helicase belonging to the superfamily 1B, which consists of 5 domains, 2 RecA like domains (1A and 2A), a 1B domain, a stalk domain, and a zinc binding domain. (Newman et al., 2021, Arya et al., 2021, Chen et al., 2022) These 5 domains are arranged in a shape of pyramid, which enables each of them to contribute to the helicase activity. (Jia et al., 2019) The zinc binding domain is located at the N-terminal and is in coordination to 3 Zn cations. (Newman et al., 2021) The 1A and 2A domains together shape a binding pocket for ATP and this accounts for the ATPase activity of NSP13. (Chen et al., 2022) Mutagenesis studies of SARS-CoV have shown that the ATP binding is relevant to residue K288, S289, D374, E375, Q404 and R507. (Jia et al., 2019) In terms of DNA duplex binding, 1A domain (residue 330-350), 2A domain (residue 516-541) and 1B domain (residue 176-186, 209-214) are involved. (Jia et al., 2019) This can be a prediction of

substrate binding in the case of SARS-CoV-2 because of the almost identical sequence between SARS-CoV and SARS-CoV-2 as mentioned above.

1.3.3 NSP13 Structure and Function Analysis

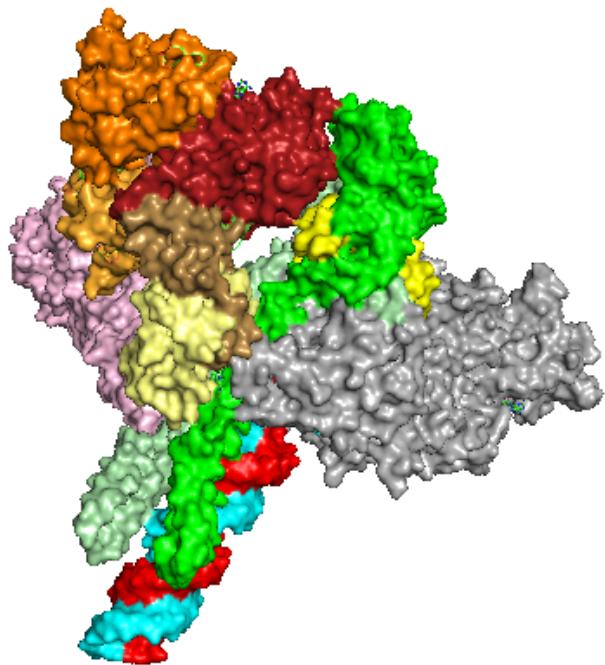


Figure 2. Structure of NSP13₂-RTC complex generated in pymol, PDB code 6xez. Colour scheme: NSP12 in grey, NSP7 in bright yellow, 2 NSP8 in bright and light green, template RNA (t-RNA) in cyan, product-RNA (p-RNA) in red, NSP13 Finger (F) in pink, NSP13 Thumb (T)-Rec1A in brick red, NSP13 (T)-Rec2A in orange, NSP13 (T)-stalk domain in brown, NSP13 (T)-1B in yellow orange (surrounded by NSP13(F), NSP13 (T)-stalk, Rec1A and 2A), NSP13 (T)-zinc binding domain in pale yellow.

The cryo-EM structure of NSP13 in RTC of SARS-CoV-2 has shown that NSP13 functions in a complex including NSP7, NSP8 and NSP12 (Figure 2) and this complex is named the NSP13₂-RTC. (Chen et al., 2022) 2 NSP are required in the assembly of NSP13₂-RTC, one is located at the thumb side of NSP12 while the other is located at the finger side (Hillen, 2021). In this thesis, these NSP13s are labeled as NSP13 (F) and NSP13 (T) respectively. The

active site of NSP13 (T) is located at the downstream of the migrating path of template-RNA (t-RNA) and the cryo-EM maps has confirmed that 5' end of t-RNA will migrate through the binding pocket of NSP13 (T). (Chen et al., 2022) However, the exact role of NSP13 (F) is still unclear and further studies are needed to confirm the helicase activity of NSP13 (F) since it may not bind to DNA or RNA judging from the structure. (Chen et al., 2020b, Malone et al., 2021, Chen et al., 2022) The structure of NSP13₂-RTC has revealed another contradiction. NSP13, as a helicase, can migrate and unwind double strand DNA or RNA from the 5' end to 3' end, but the polarity of RdRp is the opposite, which means RdRp and NSP13 migrate toward each other along the t-RNA. (Mickolajczyk et al., 2021, Saikrishnan et al., 2009, Adedeji et al., 2012, Chen et al., 2022) This is contradiction to the high replicating rate of RdRp if it translocates on t-RNA oppositely to NSP13. (Seifert et al., 2021, Chen et al., 2022) The structure analysis of NSP13₂-RTC gives an insight of additional roles of NSP13. It has been suggested that NSP13 can promote not only RNA elongation of RdRp, but also the backtracking function of RTC for proof-reading and switching t-RNA when subgenomic RNA is transcribed. (Chen et al., 2020b, Malone et al., 2021, Tavares et al., 2020, Huston et al., 2021, Chen et al., 2022) Therefore, if functions of NSP13 are inhibited, the lifecycle of the virus can be disrupted.

Four different conformational states of NSP13₂-RTC were concluded from cryo-EM maps. The largest population of NSP13₂-RTC is in a NSP13 (T)-engaged state, in which the Rec1A and 2A domain grasp the 5' end of t-RNA. (Chen et al., 2022) In this conformation, the phosphates of RNA are pointing towards Rec1A and 2A domain, while the bases are stacked in proximity to the 1B domain. This structure means the helicase can form polar interactions with the phosphate backbone and the van der Waals interactions between NSP13 and bases indicate a non-specific binding, which is different from the expectation of

specific binding to bases. When RNA does not bind to NSP13 (T), this is classified as NSP13 (T)-apo state and the Rec2A domain rotates away by 21° from Rec1A. Molecular dynamics (MD) simulations of NSP13 (T) predicted when ATP in ATP-Mg²⁺/RNA bound NSP13 (T) complex is hydrolysed, Rec2A rotates and the NSP13 (T) is in equilibrium between apo conformation and initial engaged conformation. This suggests that a stabilisation role of ATP in the NSP13 (T)-engaged conformation. 1B domain remained closed in both NSP13 (T) engaged or apo states and when 1B domain rotates by 85° with respect to stalk domain, the NSP13₂-RTC adopts a 1B-open conformation, and this is only observed with presence of NSP13 (F). MD simulations of transformation between 1B closed and open state gives an indication of a block of state transition by NSP13 (F). By overlapping simulation and cryo-EM structures, steric clash with NSP13 (F) is observed during rotation 1B and it is hypothesised that the 1B open state is caught by NSP13 (F). This transition provides an insight of the relationship between the complex structure and its functions. In the 1B open state, t-RNA approach into the RNA binding site of RdRp, and NSP13 (F) will prevent NSP13 (T) from moving opposite to RdRp, so 1B open is supposed to be RNA elongation state. In addition, the NSP13 (T) helicase activity is thought to be enhanced by NSP13 (F) by locking the 1B closed state. (Chen et al., 2022, Yan et al., 2020) The transition between 1B open and closed states can be achieved by introducing a swiveled state, in which NSP13 rotates by 38° with respect to the rest of the NSP13₂-RTC. (Chen et al., 2022) The RdRp has been shown to be able to backtrack the p-RNA and its mechanism has also been proposed from analysis of different states. (Malone et al., 2021, Chen et al., 2022) When incorrect nucleotide analog is incorporated (RNA synthesis in 1B open state), the NSP13 (T) will push the RdRp to move backward along t-RNA and produce a 3' p-RNA fragment (backtracking in NSP13 (T)-engaged state). (Chen et al., 2022) This provides a basis for proof-reading as incorrect fragment is sent out of NTP entry channel,

and endonuclease can be employed for cleavage. (Minskaia et al., 2006, Gorbalyena et al., 2006, Denison et al., 2011, Chen et al., 2022) The conformation change provides a basis of inhibition of NSP13 via allosteric sites.

1.4 Small Molecule Binding of Non-Structural Protein 13

1.4.1 Binding Pocket Search

X-ray crystal structures of NSP13 in apo, phosphate bound, and nucleotide bound state have been determined and this provides a basis for binding pocket searching. Since the phosphate bound NSP13 crystals are more robustly crystallised, by soaking them in a fragment library, 65 fragment bound structures and a number of fragments were found to be binding in pockets relevant to NSP13 functions. (Newman et al., 2021) Fifteen fragments were bound to the ATP binding site, and they are mimicking the binding interactions of ATP ribose or the base in structure. Three of the 15 fragments have a sulfonamide moiety, which makes polar interactions with the pocket, while the others interact with the pocket which is supposed to accommodate the adenine part of ATP. These 3 fragments can provide the basis for discovery of competitive inhibitors against ATP. In addition, a number of fragments were found to bind to the RNA binding channel and 3 of them bind to 2A domain, and these are residues supposed to interact with 2 phosphates in RNA-UPF1 complex. (Chakrabarti et al., 2011, Newman et al., 2021) Another fragment bound to the core of RNA binding channel with a series of polar and hydrophobic interactions while a sulfonamide analog mimicking phosphates binds to T359 and H311 in Rec1A domain and facilitated the formation of a loop from residue 337-340 by interacting with R337. (Newman et al., 2021) A binding pocket for potential RNA synthesis inhibitors was also found at the exit of p-RNA, which was suggested by 2 fragments binding to Rec1A domain and preventing RNA entry. Some fragments were thought to be allosteric inhibitors

and blocked conformational transitions of NSP13. For example, 11 fragments bound to a pocket formed by stalk domain and zinc binding domain while 2 fragments preferred a hydrophobic binding site next to Rec1A and 2A and opposite to the ATP binding site in 1B open state.

1.5 Aim of Project

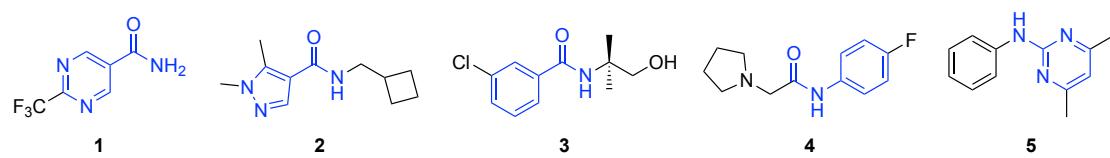
The information of the potential binding sites on NSP13 has been reported and based on the structural information and binding analysis of the literature bound fragments, a series of fragments that can potentially bind to NSP13 has been found after screening. In this project, the major task at present is to rank the screening results and find out the fragments that can have higher binding affinities than other molecules after screening. Then these higher binding score molecules need to be synthesised and purified. Characterisation of purified compounds is also needed and the ^1H , ^{13}C NMR spectra, and IR spectra will be recorded.

2 Results and discussion

2.1 Structural Based Drug Design of Non-Structural Protein 13 inhibitors

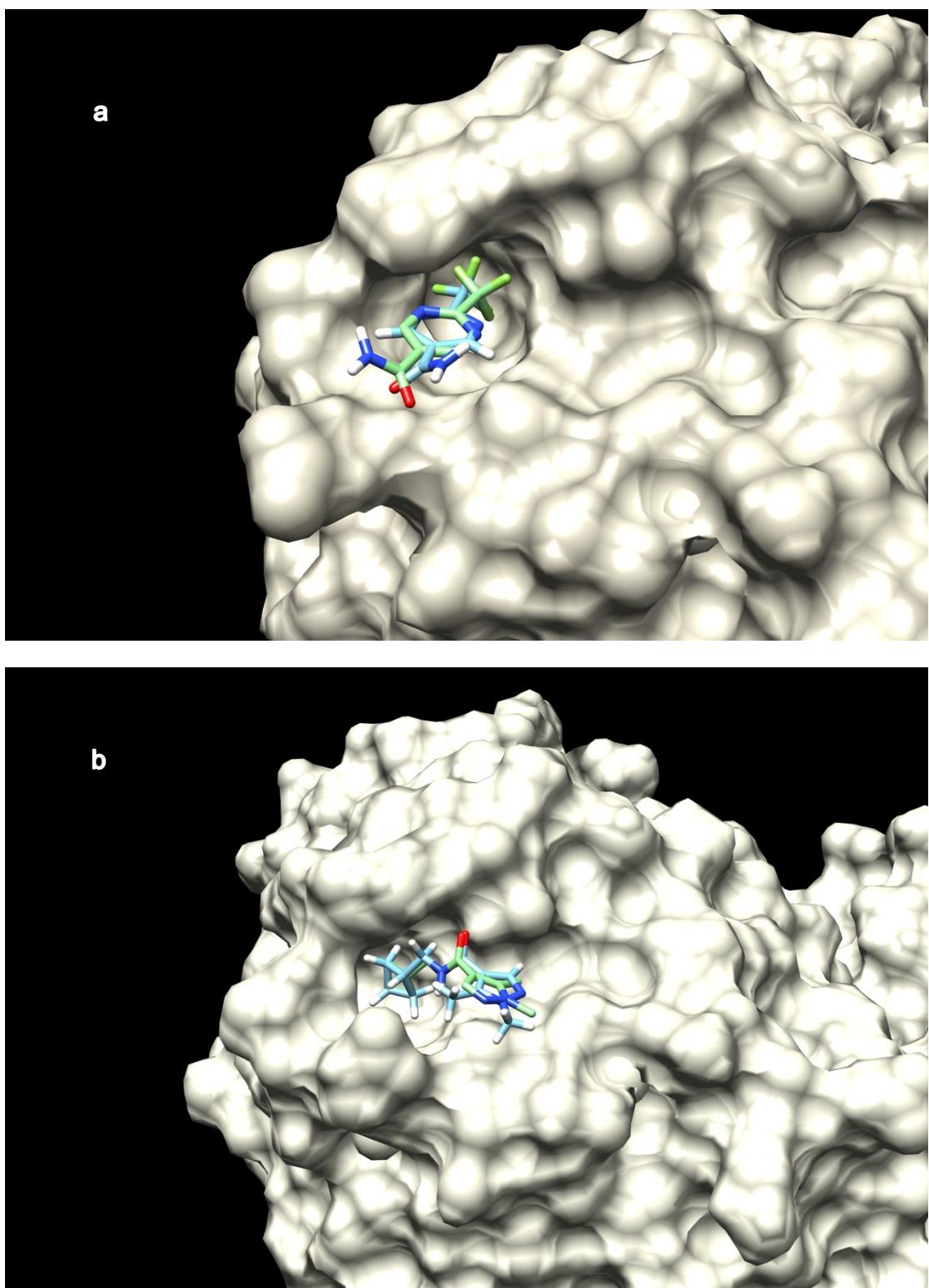
2.1.1 Binding Analysis in C-Terminus-B

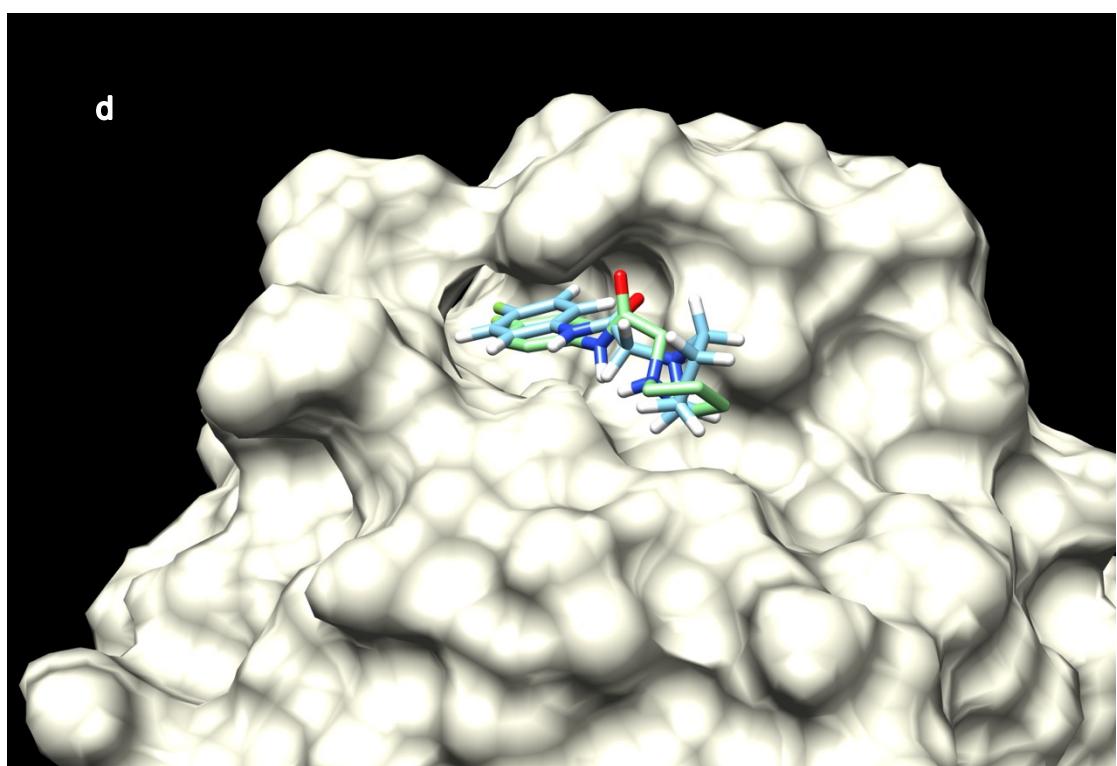
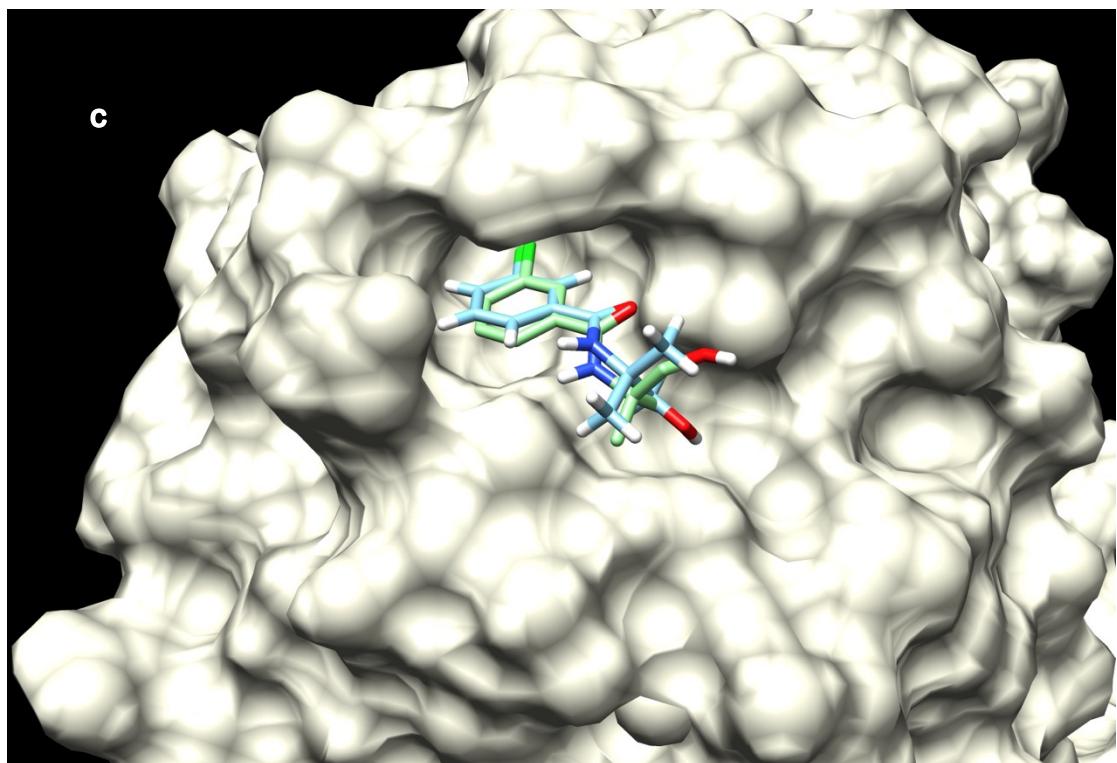
C-terminus-B is a binding site at the Rec2A domain, which is found from fragment-soaked crystal structures of NSP13. (Newman et al., 2021) The movement of Rec2A domain from the NSP13 (T)-engaged state to NSP13 (T)-apo state has been described in section 1.3.3. Therefore, this site is likely to be an allosteric site which can potentially block the conformation transition and thus inhibit the function of NSP13.



Scheme 1. Ligands reported in crystal structures by *Newman et al.* Structural similarity shown in blue

5 fragments (Scheme 1) have been shown to bind to C-terminus-B in the fragment-soaked crystal structure results. (Newman et al., 2021) Hebaalla Agha, one of the project collaborators from University North Carolina has modeled these fragment-bound structures in Pymol and finished an initial summary of a pharmacophore from these 5 fragments. 4 moieties have been suggested to be required for a decent binding of molecule to C-terminal-B. An aromatic (or aliphatic) ring substituted at meta or para to a preferentially amide or carbonyl (1 carbon away from the ring). At the other side of the amide linker, a smaller aromatic ring should be attached directly to carbonyl and there should be a N at beta position to the carbonyl. In addition, a polar group, which is linked to the smaller aromatic ring, is also needed.





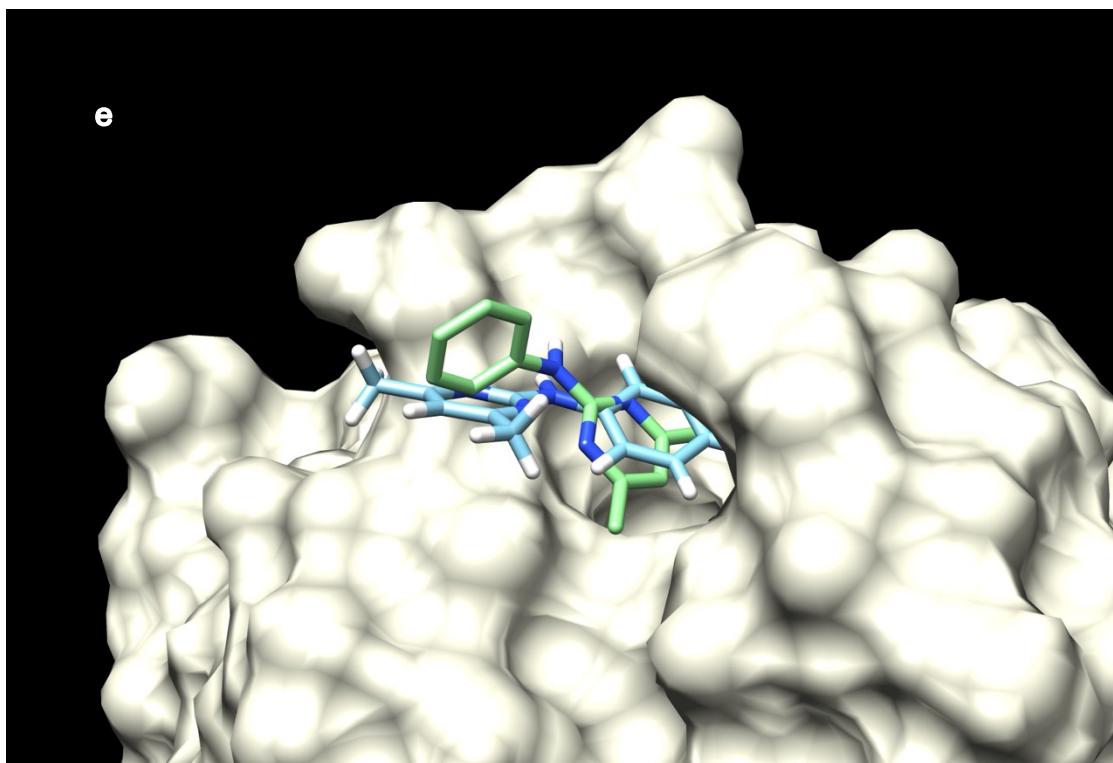


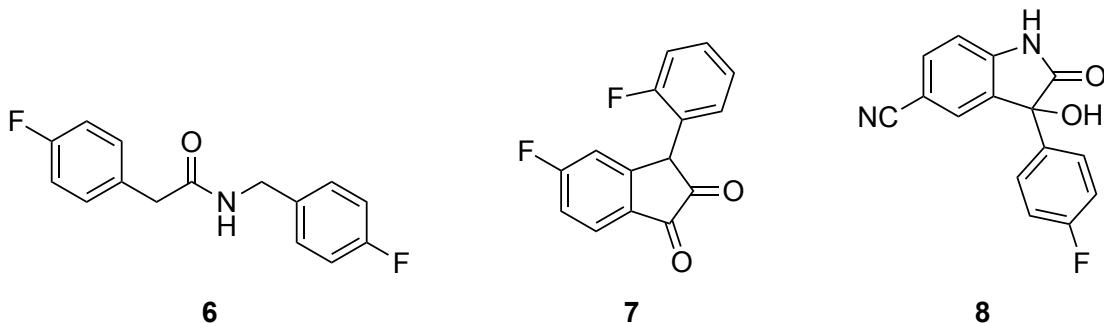
Figure 3. Overlapping of docking results (in green) and original fragment-bound crystal structure (ligand in crystal structure in cyan). Docking results generated in AutoDock Vina in Chimera. (a) molecule **1**, Vina=-4.9; (b) molecule **2**, Vina=-5.2; (c) molecule **3**, Vina=-4.9; (d) molecule **4**, Vina=-5.3; e: molecule **5**, Vina=-4.8.

The fragments reported in literature were docked in Autodock Vina. (Newman et al., 2021) The overlapping of re-docking ligand and original crystal structure is generally acceptable and the 6-membered rings in Figure 3 a-d are reasonably well-superimposed. This means Autodock Vina can predict posture of fragments relative accurately. The docking scores of literature fragments with AutoDock Vina is around -5.0 and this will be used as a reference for docking of new fragments.

2.1.2 New Fragments Predicted to Be Bound to NSP 13

As mentioned in section 1.4.2, the pharmacophore is summarised from structure features and binding interaction analysis. Based on the pharmacophore, a model was generated with fragalysis and FTMap methods

by Konstantin Popov and this model is used in screening in the Enamine REAL space library including 40 billion molecules. Approximately 500 thousand molecules are selected for docking. Meanwhile, he also generated a series of biased compounds by scoring CheEMBL library according to the pharmacophore model. After generating a list of hits, filter is applied to narrow down the number of fragments. Tom Knight suggested a filter of molecular weight of 249-269, approaching the average molecular weight of fragments in literature and logP less than 3 (rule of 3 for fragments). After docking the filtered fragments in Schrödinger Glide, 30 top-scoring hits are selected to be synthesised at this stage for the following bioassays to test the binding affinity *in vitro*. In this MSc project, 3 compounds are targeted to be synthesised as they are relatively easy to synthesise. (Scheme 2)

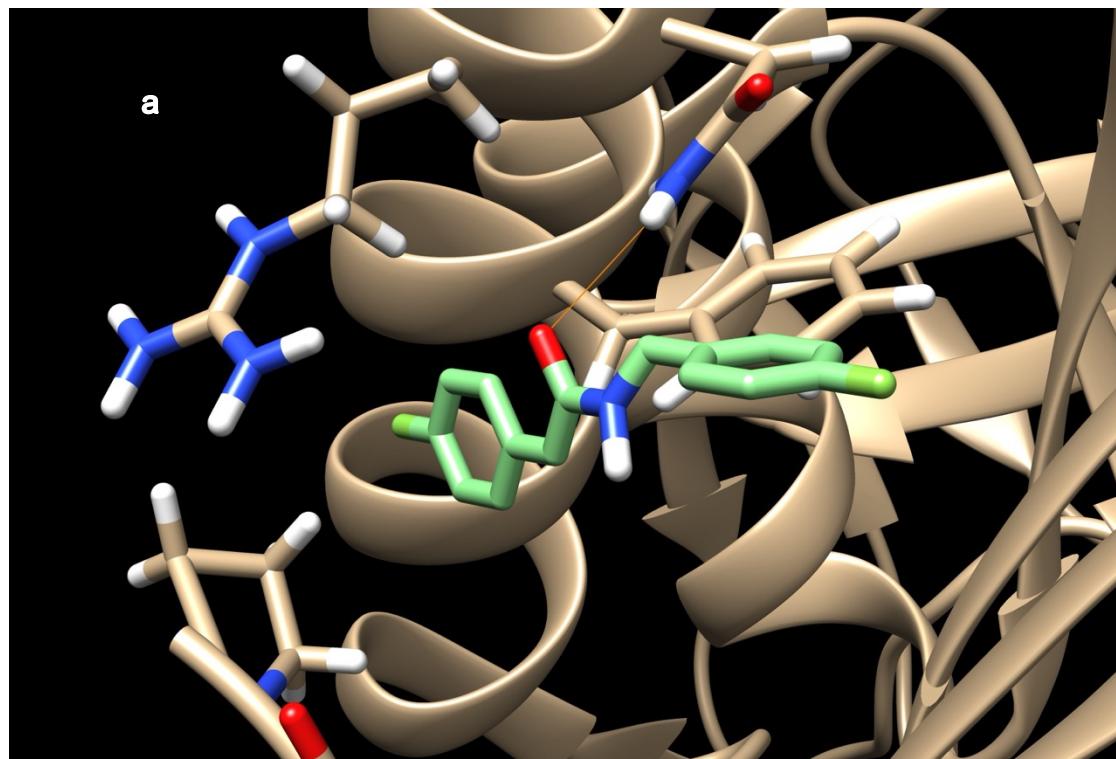


Scheme 2. Compounds targeted for synthesis in this project.

These 3 molecules are also docked in Autodock Vina to predict the binding affinity of NSP13 and illustrate the binding interactions between ligands and the protein. The Vina score of **6**, **7**, and **8** is -7.9, -8.6 and -8.7 and is more negative than that of literature ligands (-5.3 to -4.8). This means that the binding affinity of **6**, **7**, and **8** is predicted to be higher than literature fragments. However, Schrödinger Glide predicts the binding affinity in a less optimistic way. The Glide score of **6**, **7**, and **8** is -4.7, -4.6 and -4.6 respectively. The fragments bound to NSP13 literature are also docked by collaborators in contrast and the Glide score is ranging from -5.1 to -4.3. The binding affinity of the molecules **6**, **7**, and **8** is within the range of literature compounds, so they are predicted to be

at a similar level of potency compared with literature ligands. Although Glide scores do not support a better binding of **6**, **7**, and **8** and may be contradictory to Vina scores, both types of virtual docking scores predict that **6**, **7**, and **8** at least can bind to NSP13.

2.1.3 Autodock Vina Docking results



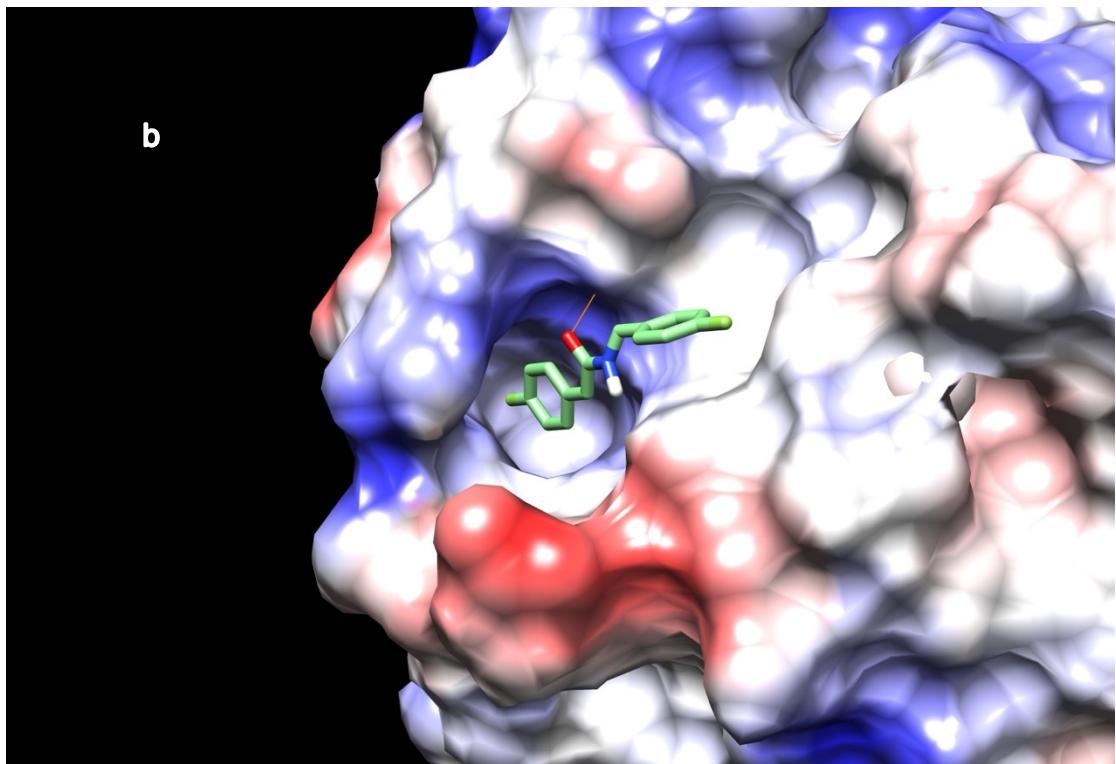


Figure 4. Docking results generated with Autodock Vina of **6**, Vina=7.9. (a) Hydrogen bonding between carbonyl O in **6** and N503 side chain amide N-H. (b) Columbic potential surface of C-terminus-B, positive region in blue, negative region in red.

The docking plot can provide insight of the interactions between **6** and NSP13. It predicts that the O of carbonyl in **6** forms a hydrogen bond with the N-H of Asn503 (Figure 4a). The phenyl ring connected to the amide N-H of **6** can potentially interact with the phenyl ring of Phe499 via $\pi - \pi$ stacking because it is closer to Phe499. The Columbic potential surface of NSP13 (Figure 4b) illustrates that the binding pocket has a positive potential, and the electronegative F preferentially points towards the pocket because it can be attracted by the binding pocket. F can attract the electron density of **6** and it has partial negative charge. This partial charge can form Coulomb interaction with electropositive binding pocket and this positive potential is likely due to positive charge of guanidium group on side chain of Arg502.

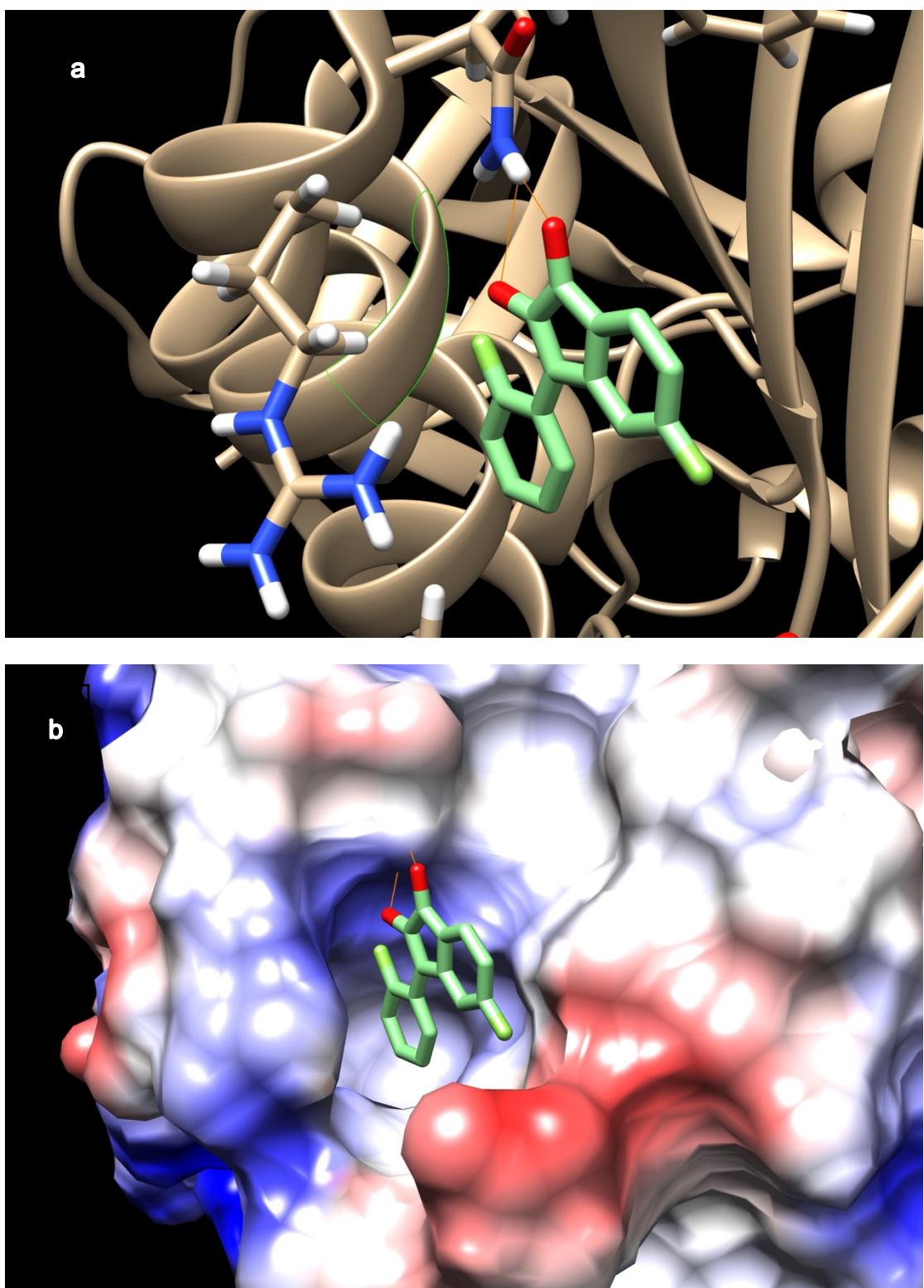
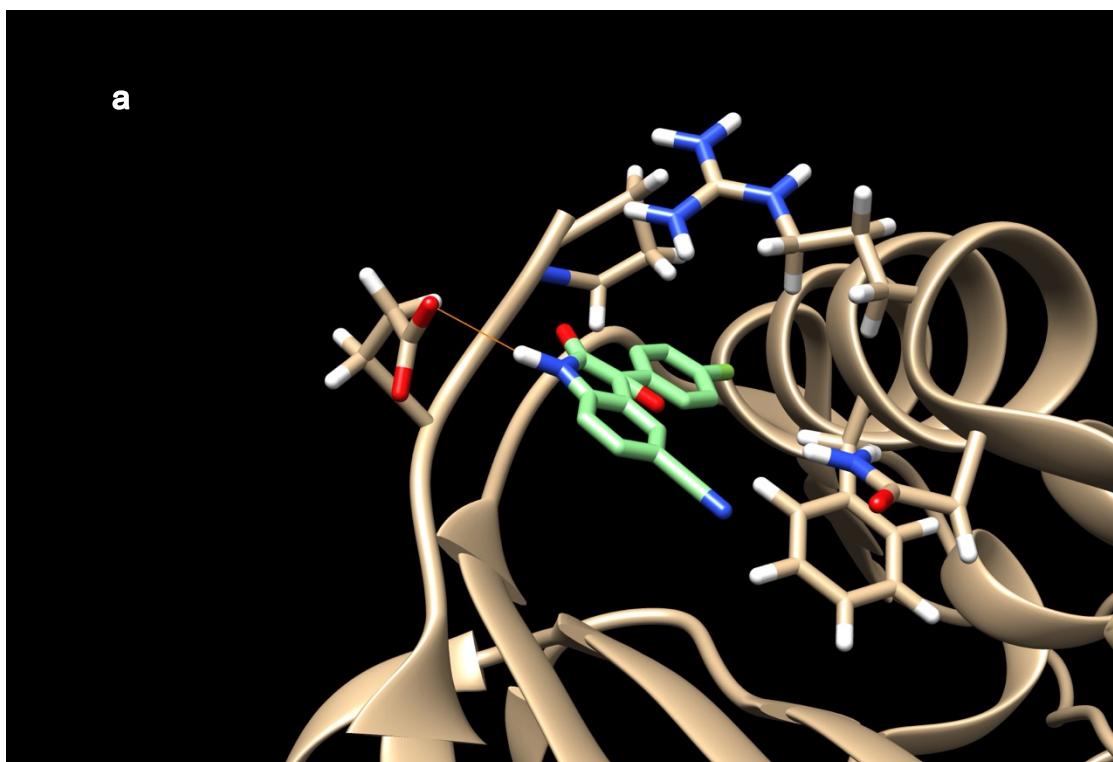


Figure 5. Docking results generated with Autodock Vina of **7**, Vina=-8.6. (a) Hydrogen bonding between carbonyl O in **7** and N503 side chain amide N-H. (b) Columbic potential surface of C-terminus-B, positive region in blue, negative region in red.

As is shown in the docking plot (Figure 5a), 2 hydrogen bonds are formed

between the 2 carbonyl O atoms and N-H in residue Asn503 like in the case of docking of **6**. Similarly, a $\pi - \pi$ stacking between the aromatic rings of Phe499 and molecule **7** is likely to be formed as well. One of fluorine atoms on **7** forms Coulomb interactions with the positive potential binding pocket inside similar to the case of **6** (Figure 5b), but the other fluorine is repelled by the negative potential surface outside. These ‘pull and push’ interactions may potentially add up and trap the fragment **7** inside the binding pocket so the docking score is more negative for **7** than **6** and potency of **7** is likely higher.



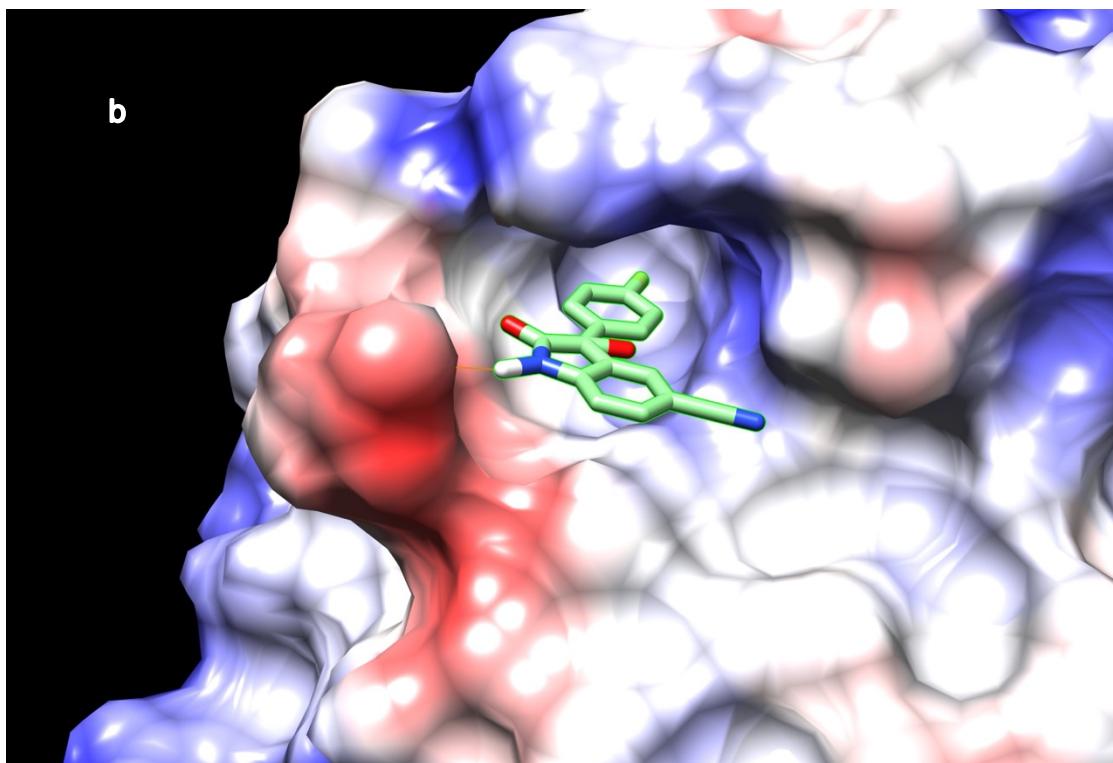
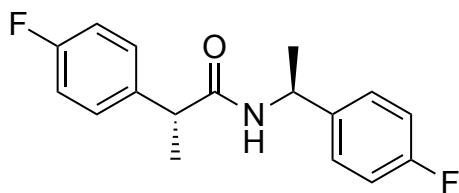


Figure 6. Docking results generated with Autodock Vina of **8**, Vina=-8.7. (a) Hydrogen bonding between amide N-H in **8** and E591 side chain O. (b) Columbic potential surface of C-terminus-B, positive region in blue, negative region in red.

The hydrogen bonds are formed with a different residue in the docking of **8** (Figure 6a). The amine N-H interacts with carboxylic acid O atoms on the sidechain of Glu591. The possibility of $\pi - \pi$ stacking between aromatic rings in **8** and Phe499 is lower than **6** and **7** because distance between the phenyl ring of Phe499 and **8** is longer. The ‘pull and push’ Coulomb interactions is also observed between **8** and the protein (Figure 6b). The F with partial negative charge is attracted to the positively charged pocket while the amide moiety in **8** faces a repulsion from the negatively charged residues like Glu591. The cyano group is also attracted to the positive potential binding pocket.

2.2 Synthesis of Top Score Fragments in Initial Screening

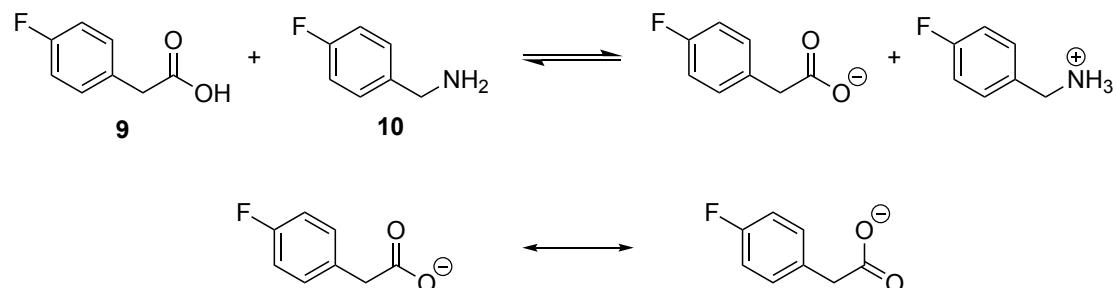
2.2.1 Synthesis of 4-Fluoro-N-[(4-fluorophenyl)methyl]benzenacetamide **6**



19

Scheme 3. Structure of (R)-2-(4-fluorophenyl)-N-((S)-1-(4-fluorophenyl)ethyl)propenamide **19**, a molecule similar to **6**.

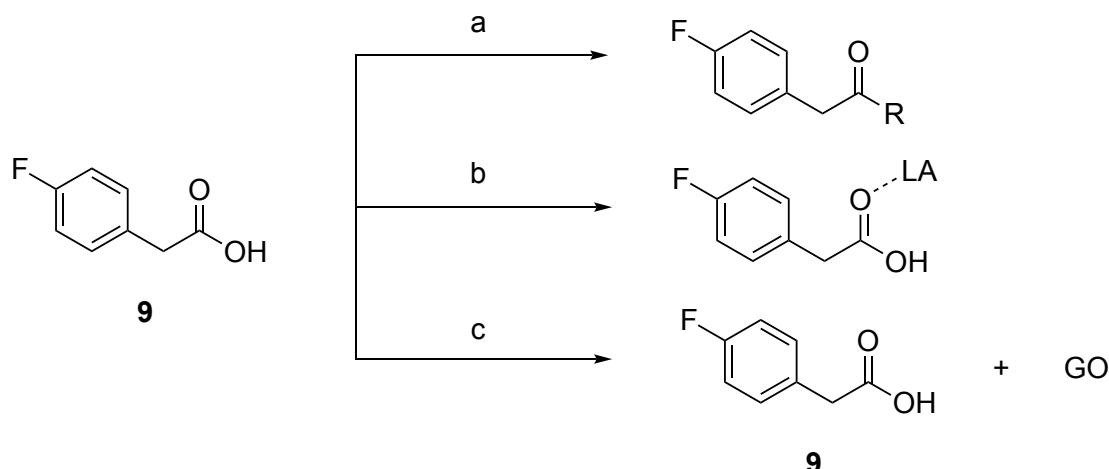
Synthesis of 4-fluoro-N-[(4-fluorophenyl)methyl]benzeneacetamide **6** has not been reported in the literature, but synthesis of a similar molecule, (R)-2-(4-fluorophenyl)-N-((S)-1-(4-fluorophenyl)ethyl)propenamide **19** was patented, including steps of Pd catalysed carbonylation and subsequent Beckmann rearrangement. Although this route is green, a high pressure is applied in the reaction (20 atm) and it is risky. Instead, the amide bond indicates that the molecule can be generated by coupling of corresponding 2-(4-fluorophenyl)acetic acid **9** and (4-fluorophenyl)methanamine **10**.



Scheme 4. Reversible proton transfer between 2-(4-fluorophenyl)acetic acid **9** and (4-fluorophenyl)methanamine **10** and resonance of 2-(4-fluorophenyl)acetic acid **9** after deprotonation.

In principle, the carbonyl in carboxylic acid is not as active as the carbonyl in corresponding acid anhydride and acyl chloride and the nucleophilic attack of amine is relatively difficult to proceed. Carboxylic acid and amine can have proton transfer between them when coexist in the system, resulting in corresponding carboxylate anion and ammonium. (Scheme 4). Meanwhile, the

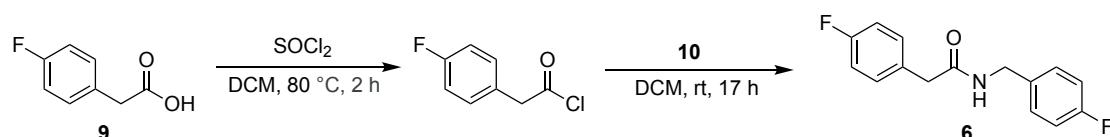
carboxylate is stabilised by resonance, and this results in more difficult nucleophilic attack. Some studies suggest that successful direct amine attack at carbonyl in carboxylic acid are only achieved at harsh thermal conditions.(Perrecux et al., 2002, Wang et al., 2008, Azizoglu et al., 2016, Gooßen et al., 2008, Charville et al., 2011, Allen et al., 2012)



Scheme 5. Synthesis options starting with 2-(4-fluorophenyl)acetic acid **9**, a: transforming **9** into more reactive species; b:using Lewis acid catalysts, LA=Lewis acid; c: using graphene oxide.

Since the direct amidation is difficult, the activation of the carbonyl is a more feasible method that renders the amidation to proceed in a relatively mild condition. At present, 3 synthesis options can be selected from (Scheme 5). The first is to use coupling agents such as triazole derivatives, carbodiimide derivatives or chlorinating agents like thionyl chloride. (Wang et al., 2018b, Chen et al., 2020a, Parlow et al., 2003, Liu and Molinski, 2011) Another option is to use Lewis acid catalysts including metals and boron species. (Lanigan et al., 2013, Mohy El Dine et al., 2015, de Azambuja and Parac-Vogt, 2019, Li et al., 2018, Anxionnat et al., 2009, Wang et al., 2020, Adolfsson et al., 2012) Lone pairs on the carbonyl oxygen atom can coordinate to a Lewis acid and make the carbonyl more electrophilic. Graphene oxide is also reported as a catalyst in amidation and can make hydrogen bonds with carboxylic acid, rendering the carbonyl more electrophilic. (Patel et al., 2020) However, organo-boron species

are difficult to remove and their generation is also tedious (Wang et al., 2020), and the selection of appropriate metal catalysts in this exact reaction can be time-consuming. Although coupling agents are more popular in amide synthesis, they may require extra effort in column chromatography because an appropriate eluent system is needed for a decent separation of different components in crude.



Scheme 6. Synthesis of 4-fluoro-N-[(4-fluorophenyl)methyl]benzenacetamide **6**, rt: room temperature.

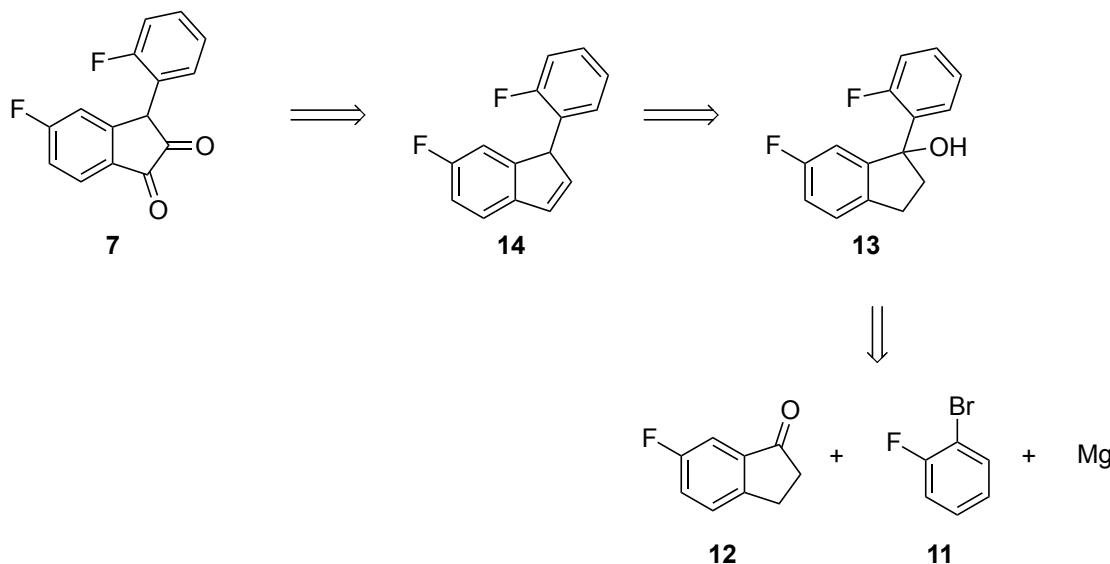
Considering of the disadvantages of coupling agents and catalysts, a transformation of carboxylic acid to acyl chloride is preferred in this synthesis. (Scheme 6.) Thionyl chloride is used for the acyl chloride formation and is volatile, so can be easily removed from the reaction system by rotary evaporation before the addition of (4-fluorophenyl)methanamine **10**. After the amine addition overnight, a yellow solution is generated and unreacted **9** was removed with saturated NaHCO_3 and dried over Na_2SO_4 anhydrous. The crude is then purified with flash column chromatography with 30% EtOAc in hexane and gives 92.4 mg of product at a 32% yield. The low yield can be explained in 2 aspects. One is that some of starting carboxylic acid **9** was not transformed into corresponding acyl chloride. The other is the acyl chloride reacted with water in the air when transferred to rotary evaporator.

The major peak in LCMS of the product was at a retention time of 3.82 min compared with 3.39 min of polar (4-fluorophenyl)methanamine **10**, which matched the moderate non-polar nature of 4-fluoro-N-[(4-

fluorophenyl)methyl]benzeneacetamide **6**. This peak also gives a m/z of 262.1, which exactly matches the molecular mass of protonated product in positive ionization mode. In addition, NMR spectroscopy showed that the isolated product was pure (Supplementary Information **s1** and **s2**) The peak at 8.53 ppm in the ¹H NMR spectrum can be easily recognised as the amide H because of its high chemical shift. When the relative integral of this peak is normalised to 1, then the integral of other peaks matches the number of other protons expected. In the ¹³C NMR spectrum, the number of peaks observed matches the number of chemically different carbons in 4-fluoro-N-[(4-fluorophenyl)methyl]benzeneacetamide **6**. FTIR spectrum also confirmed the existence of N-H (at approximately 3300 cm⁻¹) and C=O (at approximately 1700 cm⁻¹) (Supplementary Information **s3**).

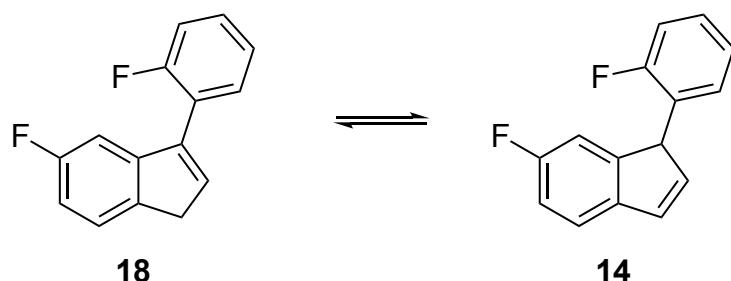
In summary, this reaction route is convenient, and the cost is relatively low. The unreacted 2-(4-fluorophenyl)acetic acid **9** is deprotonated by NaHCO₃ during the work up step and is then washed away with water. Therefore, the only impurity is excess (4-fluorophenyl)methanamine **10** (after evaporation of thionyl chloride), which can be easily removed by flash column chromatography. No reaction byproducts were observed. However, the reaction should be conducted in a water-free environment and thionyl chloride is toxic so extra care is needed when handling it.

2.2.2 Synthesis of 5-Fluoro-3-(2-fluorophenyl)-1*H*-indene-1,2(3*H*)-dione **7**



Scheme 7. Retrosynthesis of 5-fluoro-3-(2-fluorophenyl)-1*H*-indene-1,2(3*H*)-dione **7**.

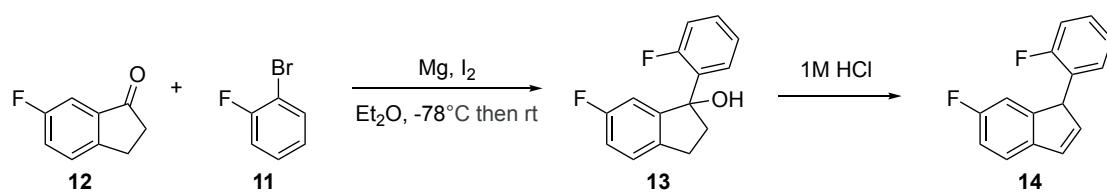
No record of direct synthesis of 5-fluoro-3-(2-fluorophenyl)-1*H*-indene-1,2(3*H*)-dione **7** has been reported. A retrosynthesis has been conducted to plan for the route to produce **7** (Scheme 7). The diketone on **7** can be formed from oxidation of an alkene and alkene can be generated by elimination. For the elimination step to proceed, a hydroxyl group should be at the 1 position of the 6-fluoroindane ring. This can account for why the fluorophenyl ring is added to the geminal position of hydroxyl because this is generated from a nucleophilic addition of Grignard reagent at the indanone carbonyl.



Scheme 8. Interconversion of post elimination product **18** and desired product **14**.

One problem should be noticed is that the elimination of hydroxyl on **13** can produce a mix of 2 different product. (Wang et al., 2021) The elimination of water can direct produce **18** and **14** can also be generated possibly by proton

transfer of post-elimination product **18** (Scheme 8). The post-elimination **18** is likely to be the major product and the fluorophenyl ring seems to prefer to attach to a sp^2 hybridised carbon. (Wang et al., 2021) There has not been any solid proof whether **18** is kinetically favoured or thermodynamically favoured. Judging from the structure of **18**, it is likely thermodynamically more stable because the bond angle of a sp^2 carbon is larger than a sp^3 carbon so sterically favourable. Although the desired product may not be favoured, enough amount of **14** can still be produced by increase the scale of reaction. A mix of products can make the purification step more difficult, but these two molecules can still be separated because their behaviors on silica should be different. Despite the drawbacks, Grignard reaction is straightforward and relatively easy to handle.



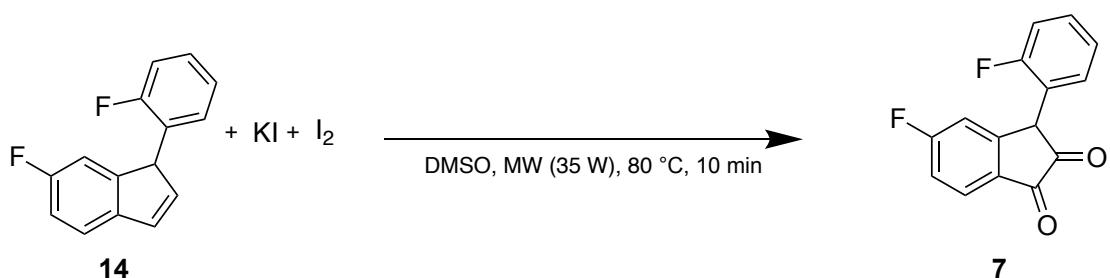
Scheme 9. Synthesis of 6-fluoro-1-(2-fluorophenyl)-1*H*-indene **14** via Grignard reaction and elimination, rt: room temperature.

The first step in the synthesis is Grignard addition to 6-fluoro-1-indanone **12** (Scheme 9). In this reaction, the Grignard reagent is first generated with 1-bromo-2-fluorobenzene **11** and Mg powder instead of Mg turnings because Mg powder is more reactive, and the generation of Grignard reagent may be easier. However, this also means that the reaction needs to start with smaller scale otherwise it can be unsafe as the generation of Grignard reagent is intensively exothermic. During the first trial, the reaction did not start after adding **11**, so a little bit of heat was applied to aid the start. After 3 h, dry ice/acetone is applied to cool the reaction down for addition of **12** and a yellow slurry can be seen after warming to room temperature. The elimination reaction may involve a protonation of hydroxyl group and water leaves E1 elimination because the tertiary carbocation is relatively stable. This trial produces a mixture of yellow

solid and oil. Thin layer chromatography (TLC) showed that 3 products were generated in the reaction and one of the spots is long wave active. **14** may not be easily protonated so the crude is examined in negative ionisation mode. LCMS provides a peak ($m/z = 227$) that matches desired product **14** or a different alkene product **18** because they have the same molecular weight. The positive ionisation mode LCMS also proved the presence of starting material **12** ($m/z = 151$). The crude was purified with flash column chromatography (ethyl acetate/hexane), but the starting material **12** was not completely removed and the separation of different products is not satisfying according to TLC results. However, the ^1H NMR spectrum of the fraction after flash column chromatography did not show a doublet due to the aliphatic proton on **14** (Supplementary Information **s4**). An analog suggests that the aliphatic proton is likely in the region of 3-4 ppm, but the spectra does not give a doublet in that region. In addition, the characteristic peaks of **14** are the alkene proton peaks but no peaks are at 4.25-6.50 ppm on the spectrum. The aliphatic region is messy, and this revealed that the post-column fraction still contains ethyl acetate and byproducts with aliphatic protons. The peak at 9.36 ppm is not related to any of protons on **11**, **12**, **13**, **14**, or **18**, so this peak may be caused by an unrecognised byproduct or the eluent for the flash column chromatography was contaminated. As is mentioned earlier, **14** may not be the major product in theory, so all products are likely unwanted byproducts and no desired **14** has been produced. The reaction was repeated with a doubled scale to make sure **14** can be produced despite it may be in small amount.

In the second trial of reaction, the generation of Grignard reagent started automatically. After adding **11**, the solution turned dark grey and many bubbles formed and this means exothermic Grignard reaction started. However, the solvent dried out after 2 h because the stopper was not totally sealed, and the dry nitrogen flow blew the solvent vapour out of the flask. The generated

Grignard reagent should not be quenched with water in the air because the whole system was under a positive pressure due to the constant dry nitrogen flow and air was less likely to enter the flask. After the whole process of reaction and work up, only yellow oil was produced this time. TLC results suggested that 3 products were produced apart from starting materials and this is consistent with the first trial of reaction. LCMS still shows a peak of $m/z = 227$ in the negative ionization mode, which matches the molecular weight of **14**, and a peak at $m/z = 151$ in positive ionization mode, indicating the presence of unreacted **12**. The crude ^1H NMR spectrum of the repetition reaction is tidier and shows a disappearance of the peak at 9.28 ppm (Supplementary Information **s5**), and this indicates that one of the byproducts is different from the crude in the last trial. Similar to the last crude ^1H NMR spectrum, the characteristic aliphatic proton doublet peak and alkene peaks are not shown. The presence of diethyl ether and ethyl acetate can also be confirmed. In addition, the 2 triplets, which give the same integral, at 2.75 and 3.11 ppm are likely from starting material because they appear in both ^1H NMR spectrum and both LCMS and TLC confirmed the presence of **11**. No aliphatic protons from byproducts are shown on the ^1H NMR spectrum. The aliphatic protons on **13** are chemically different from those on **11** so they should cause 2 more triplets in the aliphatic region. In addition, the hydroxyl group O-H is not shown on the spectrum as well. Therefore, the possibility of presence of pre-elimination intermediate **13** can be eliminated. Since the purification was not effective in the first trial of reaction, and the following product 5-fluoro-3-(2-fluorophenyl)-1H-indene-1,2(3H)-dione **7** is a highly polar diketone, which can be more easily separated, the crude was sent to the following oxidation reaction.

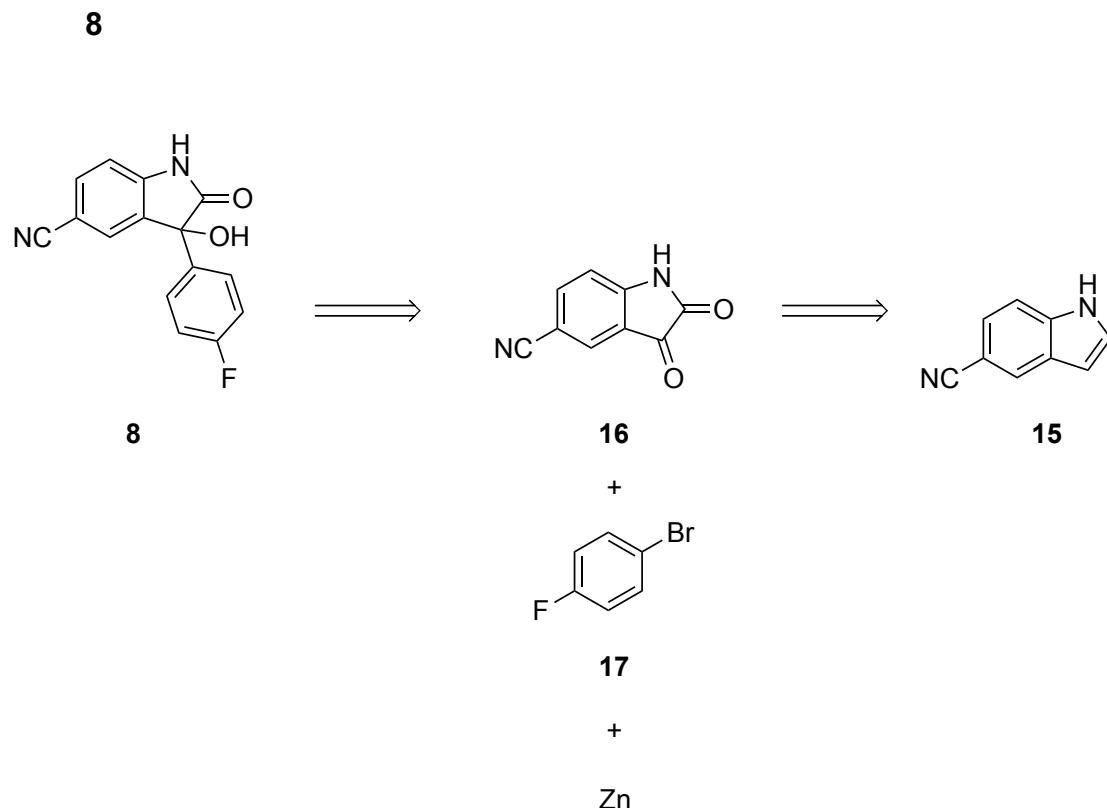


Scheme 10. Synthesis of 5-fluoro-3-(2-fluorophenyl)-1H-indene-1,2(3H)-dione **7**

The following reaction to synthesise 5-fluoro-3-(2-fluorophenyl)-1H-indene-1,2(3H)-dione **7** is microwave assisted oxidation. Without the help of microwave, the reaction condition is harsh, or catalyst is needed for the reaction. (Zeng et al., 2013, Song et al., 2020) TLC of extracted crude shows 2 new spots, one is slightly non-polar than the spot of **11**, the other is the most polar spot, which did not migrate on TLC plate in 20% EtOAc in hexane. The spot which did not migrate might be the desired product because 2 carbonyl can form strong interactions with O-H of silica. However, there was no peak at m/z = 257-259 in LCMS in either negative or positive ionisation mode. Therefore, product **7** was probably not synthesised. The absence of **7** indicate the precursor **14** may not be generated in previous steps. In order to find out whether **14** has been generated, a MeCN background was examined in LCMS and this also gives a m/z = 227 peak. Therefore, it is highly likely that **14** was not synthesised in the Grignard reaction so was absent in the starting crude. Since heat was an indication of start of Grignard reaction, it might be addition of Grignard reagent to the carbonyl did not proceed as expected. One possible reason is that the electronegative fluorine atom ortho to Mg on the Grignard reagent attract the electro density of C-Mg bond. Therefore, it may not be nucleophilic enough to attack the electrophilic carbonyl, and intermediate **13** and post-elimination product **14** were not produced. Further analysis and repetition of reaction were paused due to the time limit of the MSc project. In the future, the crude can be examined by ¹³C NMR spectroscopy to show whether carbonyls are present and how many carbonyls are present. This will be a more solid proof of absence

or presence of **7**. If it is not synthesised, the post-Grignard reaction crude needs to be purified and **14** may be separated in reverse phase columns, because the eluent has a slower gradient of polarity change. If **14** cannot be synthesised via Grignard reaction, an alternative synthesis route needs to be considered.

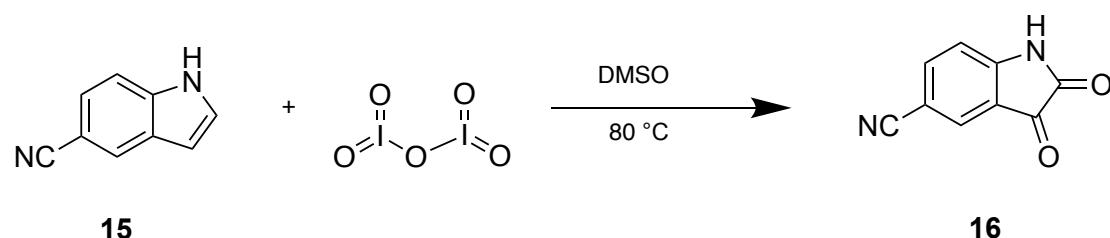
2.2.3 Synthesis of 3-(4-Fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile



Scheme 11. Retrosynthesis of 3-(4-Fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile **8**.

The retrosynthesis of 3-(4-fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile **8** is relatively straightforward. The hydroxyl and the fluorophenyl ring indicate the quaternary carbon at the 3-position of **8** is the result of nucleophilic addition at a carbonyl and the reaction is regioselective. The amide carbonyl is less reactive so the nucleophile can react with the carbonyl at 3-position preferentially. Since the Grignard reaction did not generate the pre-elimination intermediate **13** in the previous synthesis of **7**, the transfer of fluorophenyl ring was planned to be achieved via Barbier reaction. Barbier reaction is more

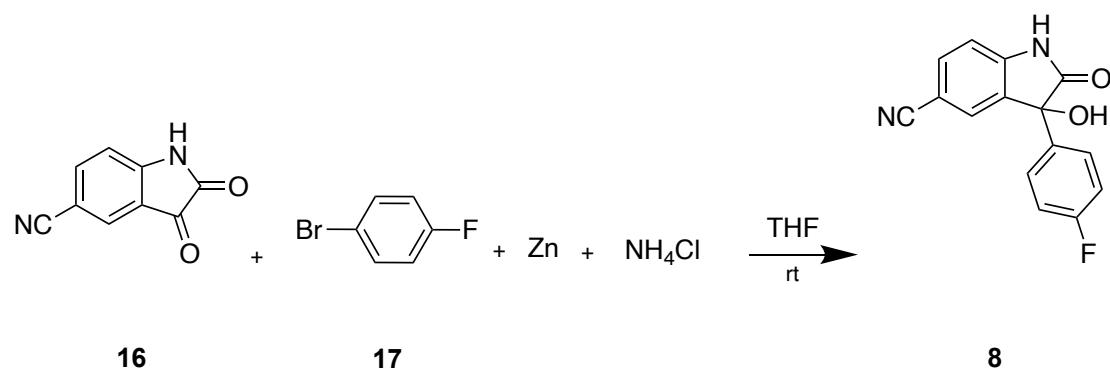
tolerant to water, so a dry nitrogen atmosphere was not required for the reaction. The structure of 2,3-dioxoindoline-5-carbonitrile **16** is similar to **7**, so **16** may also be synthesised via oxidation of an alkene.



Scheme 12. Synthesis of 2,3-dioxoindoline-5-carbonitrile **16**. DMSO: Dimethyl sulfoxide

3 synthesis methods of 2,3-dioxoindoline-5-carbonitrile **16** have been reported and all of them used hypervalent iodine species to oxidise the double bond and the reaction time with iodine pentoxide is the shortest. (Wang and Jiang, 2017, Chandra et al., 2019, Yadav et al., 2007) Dimethyl sulfoxide (DMSO) is not only a solvent but also participates in the reaction. The reaction was conducted in 3 batches and was also monitored by TLC (20% ethyl acetate in hexane). In the first batch (vial 1), the starting material **15** disappeared and 1 highly polar spot appeared in the reaction mixture. However, relatively non-polar impurities formed in the second batch (vial 2-5) and in the third batch (vial 6-8), the impurity is more non-polar than **15**. As the reaction time is the same, this may be associated with the temperature of reaction. The heating plate overheated, and temperature rose up to 90 °C and 87 °C temporarily in the second batch and third batch respectively. After extraction, a dark red crude was separated from the red fluff-like solid in the aqueous layer. The red fluff was likely the reduction product because it did not migrate on TLC plate in 50% ethyl acetate in hexane. The crude was examined with LCMS but peak at $m/z = 171\text{-}173$ can not be found in either negative or positive ionisation mode. However, ^1H NMR spectrum of the crude (Supplementary Information **s6**) indicated the existence of product **16** because of the appearance of amide N-H peaks. There are 2

peaks may be relevant to amide N-H, one at 11.43 ppm, the other at 12.04 ppm. The peak at 12.04 ppm is more deshielded so it may be the N-H on **16** and the other may be mono-oxidation byproduct with only one carbonyl. Due to the difficulty in removal of DMSO and the time limit of the project, the crude was sent for the following reaction.



Scheme 13. Synthesis of 3-(4-fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile **8**. THF: Tetrahydrofuran, rt: room temperature

The Barbier reaction used in synthesising 3-(4-fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile **8** is similar to Grignard reaction used in synthesising **14**. Both of them can transfer an aryl ring to a carbonyl. A colour change from dark red to orange was observed in the reaction. Some solid precipitated in the reaction flask and it stayed between the organic layer and aqueous layer in separation funnel. The pH of aqueous solution was found to be slightly basic. The solid seems to be less soluble in aqueous layer. The alkoxide generated after nucleophilic attack at carbonyl will take the proton in ammonium chloride solution and this leads to a basic pH of the aqueous layer. The TLC result of the crude in the first trial of reaction was interesting as the new spot did not migrate on the TLC plate in 50% ethyl acetate in hexane. Therefore, the polarity of eluent was increased (60% ethyl acetate in hexane), and the spot did migrate. The generated product seemed to be more polar than the starting material **16**. The crude LCMS confirmed the existence of product **8** as a peak matched m/z = 269 in the positive ionisation mode. However, LCMS showed a shorter

retention time of starting material **16** than **8** and this means **8** is less polar because the column of LCMS is reverse phase. Therefore, the product **8** may be overlapping with other spots on the 50% ethyl acetate TLC plate. The actual reaction time is similar to a reported analog of **8**. (Chowhan et al., 2017) The first trial of reaction proceeded for 15 min, and only 4 mg of product (3% yield) were isolated after flash column chromatography. The second trial was monitored with TLC every 20 min. 3 extra spots were shown in the first 20 min and 1 spot disappeared after 40 min. The reaction was then left overnight and only 3 spots left on TLC plate. 10 mg of product were purified with flash column chromatography (7% yield). The third trial generated 15 mg of product after column (11% yield). The yield of product is low, and this is likely because the starting material used was crude and the actual amount of **16** is less than estimated. The other reason for the low yield is production of byproducts by side reactions.

Although LCMS gives a peak matches the molecular weight of **8**, part of structural information interpreted from NMR spectra of purified **8** contradictory to LCMS. The peak at 12.09 ppm in ¹H NMR spectrum indicates the presence amide N-H, this can be consistent to the structure (Supplementary Information **s7**). The peaks at aliphatic region are caused by DMSO, water and ethyl acetate, so no alkyl protons in the compound and this is also consistent to the structure. However, the hydroxyl O-H is not seen in the spectrum, and only 4 different aromatic protons are shown in the spectrum, while 5 chemically different aromatic protons are in the structure **8**. The hydroxyl O-H may be buried in the wide water peak in the spectrum so this may account for the disappearance of O-H. The doublets of doublet at 7.51 and 7.58 ppm may be caused by the protons on the fluorophenyl ring. The proton can couple to both the adjacent proton and fluorine atom and their relative integrals are approximately the same. In addition, the coupling between adjacent protons on

the fluorophenyl ring should give the same coupling constant but the coupling constants are slightly different, 8.44 and 8.46 Hz respectively. The relative integrals of those peaks are set to 2 for reference. The peak at 7.75 ppm may be a slightly merged doublet of doublet or a quartet. The protons at 6 and 7-position of the indole-like ring can cause a doublet of doublet with a relative integral of 2. The singlet at 7.77 ppm may be caused by proton at 4-position on the indole-like ring, but its relative integral is also approximately 2. This is contradictory to the structure. More analysis has been done to investigate the identity of the compound and more inconsistency to the desired structure has been found. The distortionless enhancement by polarization transfer-90 (DEPT-90) spectrum (Supplementary Information **s7**) indicates that only 4 different aromatic protons are present in compound because it can only show CH carbons and this matches the result of ^1H NMR spectrum. A ^1H decoupled ^{19}F NMR has also been done and no peak is found in the spectrum so the fluorophenyl ring may not be added to **16**. ^1H - ^1H correlated spectroscopy (COSY) is a 2D NMR spectroscopy and is used here to investigate whether the aromatic protons are adjacent. COSY spectrum (Supplementary Information **s9**) shows that the doublets of doublet at 7.51 and 7.58 ppm in the ^1H NMR spectrum are not correlated so they are not from adjacent protons. The peaks at 7.75 and 7.77 ppm are correlated in COSY so they are correlated so they may be from protons at the indole-like ring. ^{13}C NMR spectrum (Supplementary Information **s10**) shows 7 different aromatic carbons are present while there are 10 in **8**. The carbonyl carbon is also missing in the ^{13}C NMR spectrum. If there is no carbonyl in the compound, it is contradictory to the amide N-H in ^1H NMR spectrum. The peak at 102 ppm in the ^{13}C NMR spectrum may be caused by the cyano carbon.

In the FTIR spectrum (Supplementary Information **s11**) shows the presence of amide N-H (3200 cm^{-1}), amide C=O (1600 cm^{-1}) and nitrile (2200 cm^{-1}). The hydroxyl O-H is also missing in IR spectrum, so it is likely hydroxyl is not present

in the compound. In summary, according to the analysis of the unknown compound by far, the compound seems not to be **8** and more analysis is still needed to identify the structure of the compound. Due to the time limit of the project, further analysis of the reaction is paused. If more time is given, the starting **16** should be purified for production of **8**. The Barbier reaction can also be repeated with bromobenzene in contrast. By comparing the NMR spectrum of the products from bromobenzene and **17**, it can be shown that whether the structural contradiction between spectra is caused by problems in reaction or complication by the fluorine atom.

3 Conclusion and Future Work

A series of relatively high binding scores molecules predicted in Schrödinger Glide were ranked and 30 highest score fragments were selected. 3 of 30 were targeted to be synthesised in this project. The synthesis of 4-fluoro-N-[(4-fluorophenyl)methyl]benzeneacetamide **6** was successful and the ¹H and ¹³C NMR spectra have shown the purity of **6** is high. IR spectrum also shows the amide functional group. 5-fluoro-3-(2-fluorophenyl)-1*H*-indene-1,2(3*H*)-dione **7** was not found in the reaction crude. No peak was found to match the molecular weight of **7** in LCMS. Further analysis is still needed to investigate whether the Grignard reaction proceeded in synthesis of intermediate 5-fluoro-3-(2-fluorophenyl)-1*H*-indene-1,2(3*H*)-dione **14**. Contradiction occurred in analysis of 3-(4-fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile **8**. LCMS showed a peak matching the molecular weight of **8** but ¹H, ¹³C, ¹H decoupled ¹⁹F, DEPT-90, and COSY NMR spectra have shown a different structure from **8**. In addition, ¹H NMR spectrum is also contradictory to ¹³C NMR spectrum. The hydroxyl is missing in not only ¹H NMR spectrum but also IR spectrum of **8**. The precursor **16** needs to be purified for repetition of reaction and a contrast experiment with bromobenzene can also be conducted to show whether the fluorine atom has an effect on hiding certain peaks in spectra.

In future, the synthesised fragments should be sent to examine the binding affinity to NSP13. They can be tested in surface plasmon resonance (SPR) assay. Biotinylated NSP13 will be required for the assay and they are needed to be linked on a metal surface. If the fragments are bound to NSP13, the intensity of reflected light will be changed. A saturation of response should be observed because linear response is likely the protein is binding to everything. Therefore, a contrast experiment may also be needed. As the size of NSP13 is large but the molecules are small, sensitivity may be a problem. If the SPR

assay cannot give a satisfying evaluation, isothermal titration calorimetry (ITC) assay can be an alternative option to examine the binding between fragments and NSP13. The heat emitted or absorbed will be proportional to the binding of fragments to NSP13. When the binding of synthesised fragments is examined, the second round of screening result will have been generated. Ranking the fragments from the second round of screening and synthesis of the top docking score molecules is required.

4 Experimental

Synthesis of 4-Fluoro-N-[(4-fluorophenyl)methyl]benzeneacetamide **6**

SOCl_2 (12.0 equiv) was added to 4-fluorophenylacetic acid (179 mg, 1.16 mmol) under nitrogen. After stirring for 2 h at 80 °C, volatiles removed in vacuo, and dissolved in dichloromethane (DCM) (10.0 mL), 4-fluorobenzylamine (132 μL , 1.16 mmol) was added and the mixture was stirred overnight at room temperature, then poured into saturated NaHCO_3 (25.0 mL) and extracted with DCM (3×25 mL). The combined organic layers were dried with Na_2SO_4 and concentrated in vacuo before purification with flash column chromatography (EtOAc/hexane).

Synthesis of 5-Fluoro-3-(2-fluorophenyl)-1*H*-indene-1,2(3*H*)-dione **7**

Mg powder (11.0 mmol, 0.268 g) is added to the dry reaction set up and the air is removed by nitrogen gas. A single crystal of iodine is added, and ether (10.00 mL) are then added and the mixture is stirred for 5 min. 2-bromofluorobenzene in ether solution (1.925 g, 11.00 mmol in 50 mL, 2 M) is added in small portions (5.00 mL) and the mixture is heated and refluxed for 3 h. Since the solvent dries out during this period, 2-bromofluorobenzene (0.3 mL in ether) is added. After 1h, it is cooled to -78°C with dry ice and acetone. 6-fluoro-1-indanone (1.50 g, 10 mmol) in ether (30 mL) is then added and the mixture is warmed to rt. The mixture is then poured on ice and washed with 1M HCl (100 mL), 1M NaHCO_3 (100 mL), and water (100 mL) and the organic layer is dried over MgSO_4 . The solvent is then removed in vacuo.

KI (0.250 mmol, 41.5 mg) and I₂ (0.250 mmol, 63.5 mg) are dissolved in DMSO

(2.00 mL) in a 10 mL microwave vial. The mixture is stirred for 5 min and then **14** (1.00 mmol, 228.2 mg) dissolved in DMSO (2.00 mL) is added to the vial. After sealed with Teflon septum, the mixture is heated to 80 °C and microwaved at 35 W. Extracted with Ethyl acetate (2 x 10.0 mL), washed with brine (2 x 15.0 mL), the crude is dried over Na₂SO₄. The solvent is then removed in vacuo.

Synthesis of 3-(4-Fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile **8**

1H-indole-5-carbonitrile (8x0.500 mmol, 8x71.1 mg) and iodine pentoxide (8x1.00 mmol, 8x 334 mg) are dissolved in DMSO (8x3.00 mL) and stirred at 80°C and the reaction is monitored by TLC. The reaction mixture is then washed with sat. Na₂S₂O₃ solution (8x20.0 mL), extracted with EtOAc (8x3x20.0 mL), and dried with Na₂SO₄. The solvent is removed in vacuum.

16 (0.500 mmol, 86.1 mg) and 4-bromofluorobenzene (1.50 mmol, 263 mg, 1.61 mL) are dissolved in tetrahydrofuran and Zn (2.50 mmol, 163 mg) is added and then saturated NH₄Cl solution (0.651 mL) is added. The reaction mixture is stirred and is then extracted with EtOAc (3x10.0 mL), washed with brine (5.000 mL) and dried over Na₂SO₄. The solvent is removed in vacuum. The product is purified with flash column chromatography (EtOAc/hexane).

5 Bibliography

- ADEDEJI, A. O., MARCHAND, B., TE VELTHUIS, A. J., SNIJDER, E. J., WEISS, S., EOIFF, R. L., SINGH, K. & SARAFIANOS, S. G. 2012. Mechanism of nucleic acid unwinding by SARS-CoV helicase. *PLoS One*, 7, e36521.
- ADOLFSSON, H., LUNDBERG, H. & TINNIS, F. 2012. Titanium(IV) isopropoxide as an efficient catalyst for direct amidation of nonactivated carboxylic acids. *Synlett*, 23, 2201-2204.
- AGOSTINI, M. L., ANDRES, E. L., SIMS, A. C., GRAHAM, R. L., SHEAHAN, T. P., LU, X., SMITH, E. C., CASE, J. B., FENG, J. Y., JORDAN, R., RAY, A. S., CIHLAR, T., SIEGEL, D., MACKMAN, R. L., CLARKE, M. O., BARIC, R. S. & DENISON, M. R. 2018. Coronavirus susceptibility to the antiviral Remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. *mBio*, 9.
- ALANAGREH, L., ALZOUGHOOL, F. & ATOUM, M. 2020. The human coronavirus disease COVID-19: its origin, characteristics, and insights into potential drugs and its mechanisms. *Pathogens*, 9.
- ALLEN, C. L., CHHATWAL, A. R. & WILLIAMS, J. M. 2012. Direct amide formation from unactivated carboxylic acids and amines. *Chem Commun (Camb)*, 48, 666-8.
- ANGELINI, M. M., AKHLAGHPOUR, M., NEUMAN, B. W. & BUCHMEIER, M. J. 2013. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *mBio*, 4.
- ANXIONNAT, B., GUÉRINOT, A., REYMOND, S. & COSSY, J. 2009. FeCl₃-catalyzed Ritter reaction. Synthesis of amides. *Tetrahedron Lett*, 50, 3470-3473.
- ARYA, R., KUMARI, S., PANDEY, B., MISTRY, H., BIHANI, S. C., DAS, A., PRASHAR, V., GUPTA, G. D., PANICKER, L. & KUMAR, M. 2021. Structural insights into SARS-CoV-2 proteins. *J Mol Biol*, 433, 166725.
- AZIZOGLU, S. S., KAPLAN, S., OZTURK, G., TURGUT, Y. & TOGRUL, M. 2016. Microwave-assisted synthesis of novel amide based chiral tripodal receptors for recognition of cation and anion sensors. *Arkivoc*, 2016, 44-58.
- CAVASOTTO, C. N., LAMAS, M. S. & MAGGINI, J. 2021. Functional and druggability analysis of the SARS-CoV-2 proteome. *Eur J Pharmacol*, 890, 173705.
- CHAKRABARTI, S., JAYACHANDRAN, U., BONNEAU, F., FIORINI, F., BASQUIN, C., DOMCKE, S., LE HIR, H. & CONTI, E. 2011. Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol Cell*, 41, 693-703.
- CHANDRA, A., YADAV, N. R. & MOORTHY, J. N. 2019. Facile synthesis of isatins by direct oxidation of indoles and 3-iodoindoles using NIS/IBX. *Tetrahedron*, 75, 2169-2174.
- CHARVILLE, H., JACKSON, D. A., HODGES, G., WHITING, A. & WILSON, M. R. 2011. The uncatalyzed direct amide formation reaction - mechanism studies

- and the key role of carboxylic acid H-bonding. *Eur J Org Chem*, 2011, 5981-5990.
- CHEN, G., CAO, J., WANG, Q. & ZHU, J. 2020a. Desymmetrization of prochiral cyclopentenes enabled by enantioselective palladium-catalyzed oxidative Heck reaction. *Org Lett*, 22, 322-325.
- CHEN, J., MALONE, B., LLEWELLYN, E., GRASSO, M., SHELTON, P. M. M., OLINARES, P. D. B., MARUTHI, K., ENG, E. T., VATANDASLAR, H., CHAIT, B. T., KAPOOR, T. M., DARST, S. A. & CAMPBELL, E. A. 2020b. Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell*, 182, 1560-1573 e13.
- CHEN, J., WANG, Q., MALONE, B., LLEWELLYN, E., PECHERSKY, Y., MARUTHI, K., ENG, E. T., PERRY, J. K., CAMPBELL, E. A., SHAW, D. E. & DARST, S. A. 2022. Ensemble cryo-EM reveals conformational states of the nsp13 helicase in the SARS-CoV-2 helicase replication-transcription complex. *Nat Struct Mol Biol*, 29, 250-260.
- CHEN, Y., LIU, Q. & GUO, D. 2020c. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol*, 92, 418-423.
- CHOWHAN, L. R., REDDY, M. S. & KUMAR, N. S. 2017. An efficient and rapid synthesis of 3-hydroxy-3-alkyl-2-oxindoles via Zn-mediated barbier-type reaction under aqueous conditions. *Journal of Chemical Sciences*, 129, 1205-1209.
- CUI, L., WANG, H., JI, Y., YANG, J., XU, S., HUANG, X., WANG, Z., QIN, L., TIEN, P., ZHOU, X., GUO, D. & CHEN, Y. 2015. The nucleocapsid protein of coronaviruses acts as a viral suppressor of RNA silencing in mammalian cells. *J Virol*, 89, 9029-43.
- DE AZAMBUJA, F. & PARAC-VOGT, T. N. 2019. Water-tolerant and atom economical amide bond formation by metal-substituted polyoxometalate catalysts. *ACS Catal*, 9, 10245-10252.
- DEDIEGO, M. L., ALVAREZ, E., ALMAZAN, F., REJAS, M. T., LAMIRANDE, E., ROBERTS, A., SHIEH, W. J., ZAKI, S. R., SUBBARAO, K. & ENJUANES, L. 2007. A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo. *J Virol*, 81, 1701-13.
- DEMOGINES, A., FARZAN, M. & SAWYER, S. L. 2012. Evidence for ACE2-utilizing coronaviruses (CoVs) related to severe acute respiratory syndrome CoV in bats. *J Virol*, 86, 6350-3.
- DENISON, M. R., GRAHAM, R. L., DONALDSON, E. F., ECKERLE, L. D. & BARIC, R. S. 2011. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol*, 8, 270-9.
- DOUEDI, S. & MISKOFF, J. 2020. Novel coronavirus 2019 (COVID-19): A case report and review of treatments. *Medicine (Baltimore)*, 99, e20207.
- DROZDZAL, S., ROSIK, J., LECHOWICZ, K., MACHAJ, F., SZOSTAK, B., PRZYBYCINSKI, J., LORZADEH, S., KOTFIS, K., GHAVAMI, S. & LOS, M. J. 2021. An update on drugs with therapeutic potential for SARS-CoV-2 (COVID-19) treatment. *Drug Resist Updat*, 59, 100794.

- FDA. 2022. *Know your treatment options for COVID-19* [Online]. Available: <https://www.fda.gov/consumers/consumer-updates/know-your-treatment-options-covid-19> [Accessed 5 Aug 2022].
- FRAGOULIS, G. E., MCINNES, I. B. & SIEBERT, S. 2019. JAK-inhibitors. New players in the field of immune-mediated diseases, beyond rheumatoid arthritis. *Rheumatology (Oxford)*, 58, i43-i54.
- GHORESCHI, K., LAURENCE, A. & O'SHEA, J. J. 2009. Janus kinases in immune cell signaling. *Immunol Rev*, 228, 273-87.
- GOOSSE, L., OHLMANN, D. & LANGE, P. 2008. The thermal amidation of carboxylic acids revisited. *Synth*, 2009, 160-164.
- GORBALENYA, A. E., ENJUANES, L., ZIEBUHR, J. & SNIJDER, E. J. 2006. Nidovirales: evolving the largest RNA virus genome. *Virus Res*, 117, 17-37.
- GORDON, C. J., TCHESENOKOV, E. P., FENG, J. Y., PORTER, D. P. & GOTTE, M. 2020a. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *J Biol Chem*, 295, 4773-4779.
- GORDON, C. J., TCHESENOKOV, E. P., WOOLNER, E., PERRY, J. K., FENG, J. Y., PORTER, D. P. & GOTTE, M. 2020b. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J Biol Chem*, 295, 6785-6797.
- HABTEMARIAM, S., NABAVI, S. F., BANACH, M., BERINDAN-NEAGOE, I., SARKAR, K., SIL, P. C. & NABAVI, S. M. 2020. Should we try SARS-CoV-2 helicase inhibitors for COVID-19 therapy? *Arch Med Res*, 51, 733-735.
- HAMMING, I., TIMENS, W., BULTHUIS, M. L., LELY, A. T., NAVIS, G. & VAN GOOR, H. 2004. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol*, 203, 631-7.
- HELMY, Y. A., FAWZY, M., ELASWAD, A., SOBIEH, A., KENNEY, S. P. & SHEHATA, A. A. 2020. The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *J Clin Med*, 9.
- HILLEN, H. S. 2021. Structure and function of SARS-CoV-2 polymerase. *Curr Opin Virol*, 48, 82-90.
- HOFFMANN, M., KLEINE-WEBER, H., SCHROEDER, S., KRUGER, N., HERRLER, T., ERICHSEN, S., SCHIERGENS, T. S., HERRLER, G., WU, N. H., NITSCHE, A., MULLER, M. A., DROSTEN, C. & POHLMANN, S. 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181, 271-280 e8.
- HUANG, C., WANG, Y., LI, X., REN, L., ZHAO, J., HU, Y., ZHANG, L., FAN, G., XU, J., GU, X., CHENG, Z., YU, T., XIA, J., WEI, Y., WU, W., XIE, X., YIN, W., LI, H., LIU, M., XIAO, Y., GAO, H., GUO, L., XIE, J., WANG, G., JIANG, R., GAO, Z., JIN, Q., WANG, J. & CAO, B. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, 395, 497-506.

- HUSTON, N. C., WAN, H., STRINE, M. S., DE CESARIS ARAUJO TAVARES, R., WILEN, C. B. & PYLE, A. M. 2021. Comprehensive *in vivo* secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol Cell*, 81, 584-598 e5.
- IMBERT, I., SNIJDER, E. J., DIMITROVA, M., GUILLEMOT, J. C., LECINE, P. & CANARD, B. 2008. The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. *Virus Res*, 133, 136-48.
- JIA, H. P., LOOK, D. C., SHI, L., HICKEY, M., PEWE, L., NETLAND, J., FARZAN, M., WOHLFORD-LENANE, C., PERLMAN, S. & MCCRAY, P. B., JR. 2005. ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. *J Virol*, 79, 14614-21.
- JIA, Z., YAN, L., REN, Z., WU, L., WANG, J., GUO, J., ZHENG, L., MING, Z., ZHANG, L., LOU, Z. & RAO, Z. 2019. Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res*, 47, 6538-6550.
- JORGENSEN, S. C. J., KEBRIAEI, R. & DRESSER, L. D. 2020a. Remdesivir: Review of pharmacology, pre-clinical data, and emerging clinical experience for COVID-19. *Pharmacotherapy*, 40, 659-671.
- JORGENSEN, S. C. J., TSE, C. L. Y., BURRY, L. & DRESSER, L. D. 2020b. Baricitinib: A review of pharmacology, safety, and emerging clinical experience in COVID-19. *Pharmacotherapy*, 40, 843-856.
- KAMITANI, W., HUANG, C., NARAYANAN, K., LOKUGAMAGE, K. G. & MAKINO, S. 2009. A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nat Struct Mol Biol*, 16, 1134-40.
- KLEYMANN, G., FISCHER, R., BETZ, U. A. K., HENDRIX, M., BENDER, W., SCHNEIDER, U., HANDKE, G., ECKENBERG, P., HEWLETT, G., PEVZNER, V., BAUMEISTER, J., WEBER, O., HENNINGER, K., KELDENRICH, J., JENSEN, A., KOLB, J., BACH, U., POPP, A., MÄBEN, J., FRAPPA, I., HAEBICH, D., LOCKHOFF, O. & RÜBSAMEN-WAIGMANN, H. 2002. New helicase-primase inhibitors as drug candidates for the treatment of herpes simplex disease. *Nat Med*, 8, 392-398.
- LANIGAN, R. M., STARKOV, P. & SHEPPARD, T. D. 2013. Direct synthesis of amides from carboxylic acids and amines using B(OCH₂CF₃)₃. *J Org Chem*, 78, 4512-23.
- LEE, H. J., SHIEH, C. K., GORBALENYA, A. E., KOONIN, E. V., LAMONICA, N., TULER, J., BAGDZHADZYAN, A. & LAI, M. M. 1991. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virol*, 180, 567-582.
- LI, N., WANG, L., ZHANG, L., ZHAO, W., QIAO, J., XU, X. & LIANG, Z. 2018. Air-stable bis(pentamethylcyclopentadienyl) zirconium perfluorooctanesulfonate as an efficient and recyclable catalyst for the synthesis of N-substituted amides. *ChemCatChem*, 10, 3532-3538.

- LI, W., MOORE, M. J., VASILIEVA, N., SUI, J., WONG, S. K., BERNE, M. A., SOMASUNDARAN, M., SULLIVAN, J. L., LUZURIAGA, K., GRENNOUGH, T. K., CHOE, H. & MICHAEL, F. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426, 450-454.
- LIU, C. & MOLINSKI, T. F. 2011. Preparation of alpha-amino acids by oxidative oxazoline-oxazinone rearrangement-hydrogenation (OOOH). Scope and limitations. *Chem Asian J*, 6, 2022-7.
- LIU, K. D., GAFFEN, S. L. & GOLDSMITH, M. A. 1998. JAK/STAT signaling by cytokine receptors. *Curr Opin Immunol*, 10, 271-278.
- MALONE, B., CHEN, J., WANG, Q., LLEWELLYN, E., CHOI, Y. J., OLINARES, P. D. B., CAO, X., HERNANDEZ, C., ENG, E. T., CHAIT, B. T., SHAW, D. E., LANDICK, R., DARST, S. A. & CAMPBELL, E. A. 2021. Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex. *Proc Natl Acad Sci U S A*, 118.
- MARZI, M., VAKIL, M. K., BAHMANYAR, M. & ZARENEZHAD, E. 2022. Paxlovid: Mechanism of action, synthesis, and in silico study. *Biomed Res Int*, 2022, 7341493.
- MICKOLAJCZYK, K. J., SHELTON, P. M. M., GRASSO, M., CAO, X., WARRINGTON, S. E., AHER, A., LIU, S. & KAPOOR, T. M. 2021. Force-dependent stimulation of RNA unwinding by SARS-CoV-2 nsp13 helicase. *Biophys J*, 120, 1020-1030.
- MINSKAIA, E., HERTZIG, T., GORBALENYA, A. E., CAMPANACCI, V., CAMBILLAU, C., CANARD, B. & ZIEBUHR, J. 2006. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci U S A*, 103, 5108-13.
- MOHY EL DINE, T., ERB, W., BERHAULT, Y., ROUDEN, J. & BLANCHET, J. 2015. Catalytic chemical amide synthesis at room temperature: one more step toward peptide synthesis. *J Org Chem*, 80, 4532-44.
- NEUMAN, B. W., CHAMBERLAIN, P., BOWDEN, F. & JOSEPH, J. 2014. Atlas of coronavirus replicase structure. *Virus Res*, 194, 49-66.
- NEUMAN, B. W., KISS, G., KUNDING, A. H., BHELLA, D., BAKSH, M. F., CONNELLY, S., DROESE, B., KLAUS, J. P., MAKINO, S., SAWICKI, S. G., SIDDELL, S. G., STAMOU, D. G., WILSON, I. A., KUHN, P. & BUCHMEIER, M. J. 2011. A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol*, 174, 11-22.
- NEWMAN, J. A., DOUANGAMATH, A., YADZANI, S., YOSAATMADJA, Y., AIMON, A., BRANDAO-NETO, J., DUNNETT, L., GORRIE-STONE, T., SKYNER, R., FEARON, D., SCHAPIRA, M., VON DELFT, F. & GILEADI, O. 2021. Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. *Nat Commun*, 12, 4848.
- O'DONOUGHUE, S. I., SCHAFERHANS, A., SIKTA, N., STOLTE, C., KAUR, S., HO, B. K., ANDERSON, S., PROCTER, J. B., DALLAGO, C., BORDIN, N., ADCOCK, M. & ROST, B. 2021. SARS-CoV-2 structural coverage map

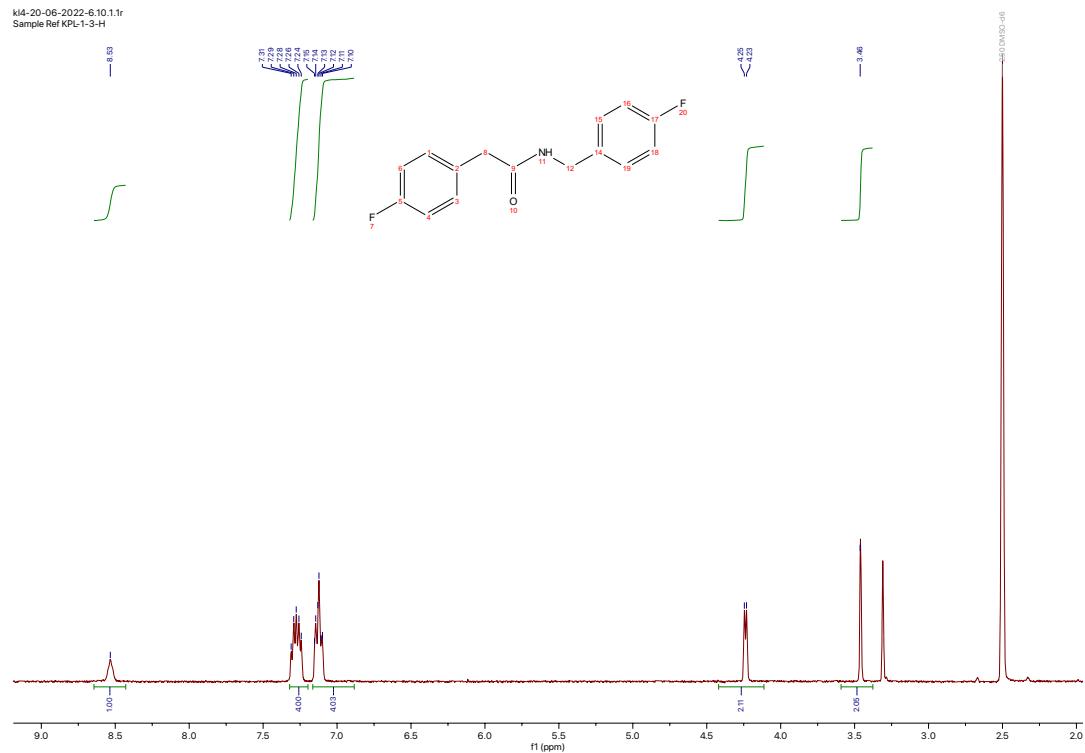
- reveals viral protein assembly, mimicry, and hijacking mechanisms. *Mol Syst Biol*, 17, e10079.
- PARLOW, J. J., STEVENS, A. M., STEGEMAN, R. A., STALLINGS, W. C., KURUMBAIL, R. G. & SOUTH, M. S. 2003. Synthesis and crystal structures of substituted benzenes and benzoquinones as tissue factor VIIa inhibitors. *J Med Chem*, 46, 4297-4312.
- PATEL, K. P., GAYAKWAD, E. M. & SHANKARLING, G. S. 2020. Graphene oxide as a metal-free carbocatalyst for direct amide synthesis from carboxylic acid and amine under solvent-free reaction condition. *ChemistrySelect*, 5, 8295-8300.
- PERRECUX, L., LOUPY, A. & VOLATRON, F. 2002. Solvent-free preparation of amides from acids and primary amines under microwave irradiation. *Tetrahedron*, 2155-2162.
- POURKARIM, F., POURTAGHI-ANVARIAN, S. & REZAEI, H. 2022. Molnupiravir: A new candidate for COVID-19 treatment. *Pharmacol Res Perspect*, 10, e00909.
- QIAN, Z., TRAVANTY, E. A., OKO, L., EDEEN, K., BERGLUND, A., WANG, J., ITO, Y., HOLMES, K. V. & MASON, R. J. 2013. Innate immune response of human alveolar type II cells infected with severe acute respiratory syndrome-coronavirus. *Am J Respir Cell Mol Biol*, 48, 742-8.
- QIN, C., ZHOU, L., HU, Z., ZHANG, S., YANG, S., TAO, Y., XIE, C., MA, K., SHANG, K., WANG, W. & TIAN, D. S. 2020. Dysregulation of immune response in patients with coronavirus 2019 (COVID-19) in Wuhan, China. *Clin Infect Dis*, 71, 762-768.
- RICHARDSON, P., GRIFFIN, I., TUCKER, C., SMITH, D., OEHHSLE, O., PHELAN, A., RAWLING, M., SAVORY, E. & STEBBING, J. 2020. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *The Lancet*, 395, e30-e31.
- ROMANO, M., RUGGIERO, A., SQUEGLIA, F., MAGA, G. & BERISIO, R. 2020. A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells*, 9.
- SAIKRISHNAN, K., POWELL, B., COOK, N. J., WEBB, M. R. & WIGLEY, D. B. 2009. Mechanistic basis of 5'-3' translocation in SF1B helicases. *Cell*, 137, 849-59.
- SEIFERT, M., BERA, S. C., VAN NIES, P., KIRCHDOERFER, R. N., SHANNON, A., LE, T. T., MENG, X., XIA, H., WOOD, J. M., HARRIS, L. D., PAPINI, F. S., ARNOLD, J. J., ALMO, S., GROVE, T. L., SHI, P. Y., XIANG, Y., CANARD, B., DEPKEN, M., CAMERON, C. E. & DULIN, D. 2021. Inhibition of SARS-CoV-2 polymerase by nucleotide analogs from a single-molecule perspective. *Elife*, 10.
- SEYBERT, A., POSTHUMA, C. C., VAN DINTEN, L. C., SNIJDER, E. J., GORBALENYA, A. E. & ZIEBUHR, J. 2005. A complex zinc finger controls the enzymatic activities of nidovirus helicases. *J Virol*, 79, 696-704.
- SHANNON, A., LE, N. T., SELISKO, B., EYDOUX, C., ALVAREZ, K., GUILLEMOT, J. C., DECROLY, E., PEERSEN, O., FERRON, F. & CANARD, B. 2020.

- Remdesivir and SARS-CoV-2: Structural requirements at both nsp12 RdRp and nsp14 Exonuclease active-sites. *Antiviral Res*, 178, 104793.
- SHUM, K. T. & TANNER, J. A. 2008. Differential inhibitory activities and stabilisation of DNA aptamers against the SARS coronavirus helicase. *Chembiochem*, 9, 3037-45.
- SINGH, A. K., SINGH, A., SINGH, R. & MISRA, A. 2020. Remdesivir in COVID-19: A critical review of pharmacology, pre-clinical and clinical studies. *Diabetes Metab Syndr*, 14, 641-648.
- SONG, T., MA, Z., REN, P., YUAN, Y., XIAO, J. & YANG, Y. 2020. A bifunctional iron nanocomposite catalyst for efficient oxidation of alkenes to ketones and 1,2-diketones. *ACS Catalysis*, 10, 4617-4629.
- TAVARES, R. C. A., MAHADESHWAR, G., WAN, H., HUSTON, N. C. & PYLE, A. M. 2020. The global and local distribution of RNA structure throughout the SARS-CoV-2 genome. *J Virol*.
- TIAN, S., HU, W., NIU, L., LIU, H., XU, H. & XIAO, S. Y. 2020. Pulmonary pathology of early-phase 2019 novel coronavirus (COVID-19) pneumonia in two patients with lung cancer. *J Thorac Oncol*, 15, 700-704.
- WANG, C.-P. & JIANG, G.-F. 2017. An efficient method based on indoles for the synthesis of isatins by taking advantage of I₂O₅ as oxidant. *Tetrahedron Lett*, 58, 1747-1750.
- WANG, H., DONG, W., HOU, Z., CHENG, L., LI, X. & HUANG, L. 2020. Direct amidation of non-activated carboxylic acid and amine derivatives catalyzed by TiCp₂Cl₂. *Appl Organomet Chem*, 34.
- WANG, M., HAO, H., LEEPER, N. J., ZHU, L. & EARLY CAREER, C. 2018a. Thrombotic regulation from the endothelial cell perspectives. *Arterioscler Thromb Vasc Biol*, 38, e90-e95.
- WANG, X., LIU, B., SEARLE, X., YEUNG, C., BOGDAN, A., GRESZLER, S., SINGH, A., FAN, Y., SWENSEN, A. M., VORTHERMS, T., BALUT, C., JIA, Y., DESINO, K., GAO, W., YONG, H., TSE, C. & KYM, P. 2018b. Discovery of 4-[(2R,4R)-4-({[1-(2,2-difluoro-1,3-benzodioxol-5-yl)cyclopropyl]carbonyl}amino)-7-(difluoromethoxy)-3,4-dihydro-2H-chromen-2-yl]benzoic Acid (ABBV/GLPG-2222), a potent cystic fibrosis transmembrane conductance regulator (CFTR) corrector for the treatment of cystic fibrosis. *J Med Chem*, 61, 1436-1449.
- WANG, X. J., YANG, Q., LIU, F. & YOU, Q. D. 2008. Microwave-assisted synthesis of amide under solvent-free conditions. *Synth Commun*, 38, 1028-1035.
- WANG, Z., LI, Y., CHEN, F., QIAN, P. C. & CHENG, J. 2021. The intramolecular reaction of acetophenone N-tosylhydrazone and vinyl: Bronsted acid-promoted cationic cyclization toward polysubstituted indenes. *Chem Commun (Camb)*, 57, 1810-1813.
- WEISS, S. R. & NAVAS-MARTIN, S. 2005. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev*, 69, 635-64.

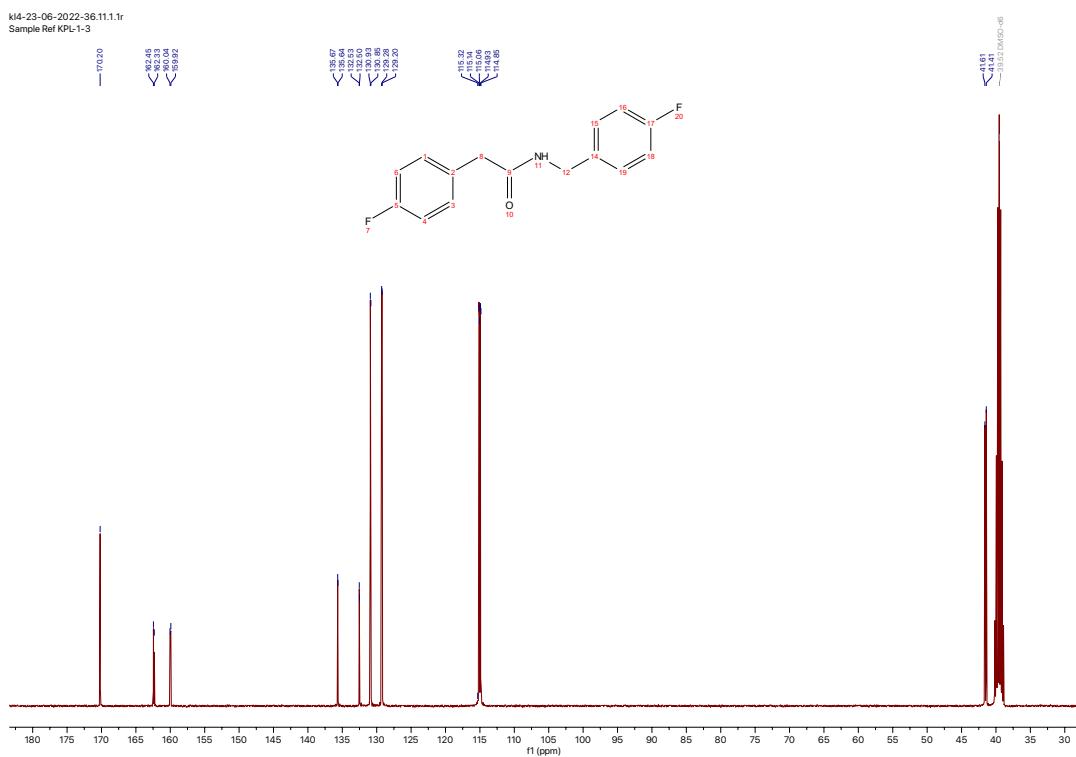
- WHO. 2022. *Coronavirus (COVID-19) data* [Online]. Available: <https://www.who.int/data> [Accessed 5 Aug 2022].
- XU, Z., SHI, L., WANG, Y., ZHANG, J., HUANG, L., ZHANG, C., LIU, S., ZHAO, P., LIU, H., ZHU, L., TAI, Y., BAI, C., GAO, T., SONG, J., XIA, P., DONG, J., ZHAO, J. & WANG, F.-S. 2020. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir Med*, 8, 420-422.
- YADAV, J., REDDY, B., REDDY, C. & KRISHNA, A. 2007. Indium(III) chloride/2-iodoxybenzoic acid: a novel reagent system for the conversion of indoles into isatins. *Synthesis*, 2007, 693-696.
- YAN, L., ZHANG, Y., GE, J., ZHENG, L., GAO, Y., WANG, T., JIA, Z., WANG, H., HUANG, Y., LI, M., WANG, Q., RAO, Z. & LOU, Z. 2020. Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nat Commun*, 11, 5874.
- YUKI, K., FUJIOGI, M. & KOUTSOGIANNAKI, S. 2020. COVID-19 pathophysiology: A review. *Clin Immunol*, 215, 108427.
- ZENG, H., PAPPAS, C., BELSER, J. A., HOUSER, K. V., ZHONG, W., WADFORD, D. A., STEVENS, T., BALCZON, R., KATZ, J. M. & TUMPEY, T. M. 2012. Human pulmonary microvascular endothelial cells support productive replication of highly pathogenic avian influenza viruses: possible involvement in the pathogenesis of human H5N1 virus infection. *J Virol*, 86, 667-78.
- ZENG, X., MIAO, C., WANG, S., XIA, C. & SUN, W. 2013. Facile and highly chemoselective synthesis of benzil derivatives via oxidation of stilbenes in an I₂-H₂O system. *RSC Advances*, 3.
- ZHENG, M., GAO, Y., WANG, G., SONG, G., LIU, S., SUN, D., XU, Y. & TIAN, Z. 2020. Functional exhaustion of antiviral lymphocytes in COVID-19 patients. *Cell Mol Immunol*, 17, 533-535.
- ZHOU, Y., FU, B., ZHENG, X., WANG, D., ZHAO, C., QI, Y., SUN, R., TIAN, Z., XU, X. & WEI, H. 2020. Pathogenic T-cells and inflammatory monocytes incite inflammatory storms in severe COVID-19 patients. *Natl Sci Rev*, 7, 998-1002.
- ZHU, N., ZHANG, D., WANG, W., LI, X., YANG, B., SONG, J., ZHAO, X., HUANG, B., SHI, W., LU, R., NIU, P., ZHAN, F., MA, X., WANG, D., XU, W., WU, G., GAO, G. F., TAN, W., CHINA NOVEL CORONAVIRUS, I. & RESEARCH, T. 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*, 382, 727-733.
- ZIEBUHR, J., SNIJDER, E. J. & GORBALENYA, A. E. 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J Gen Virol*, 81, 853-879.

6 Appendix

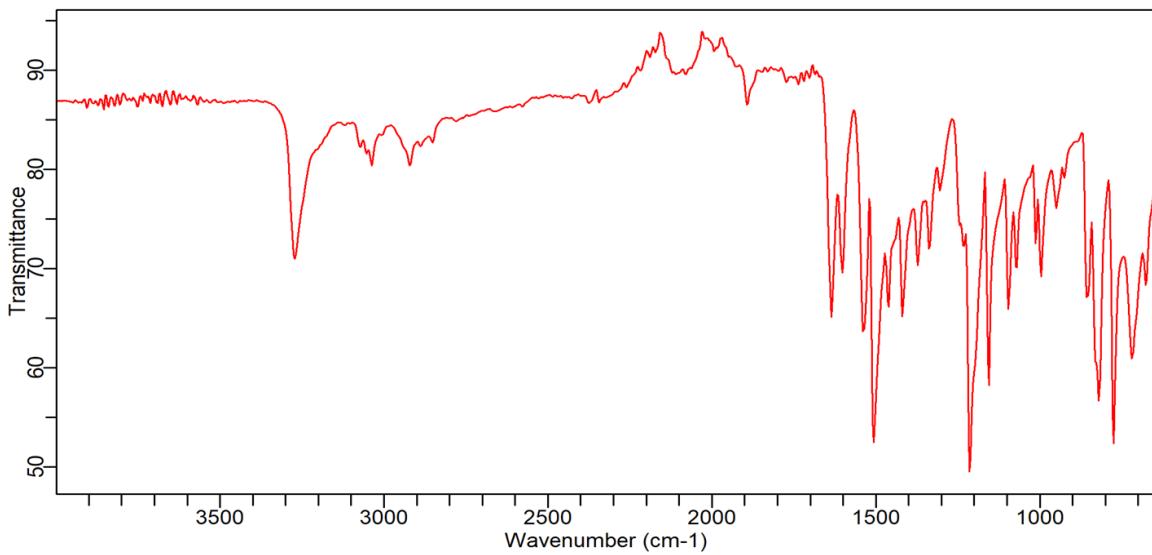
6.1 Supplementary Information



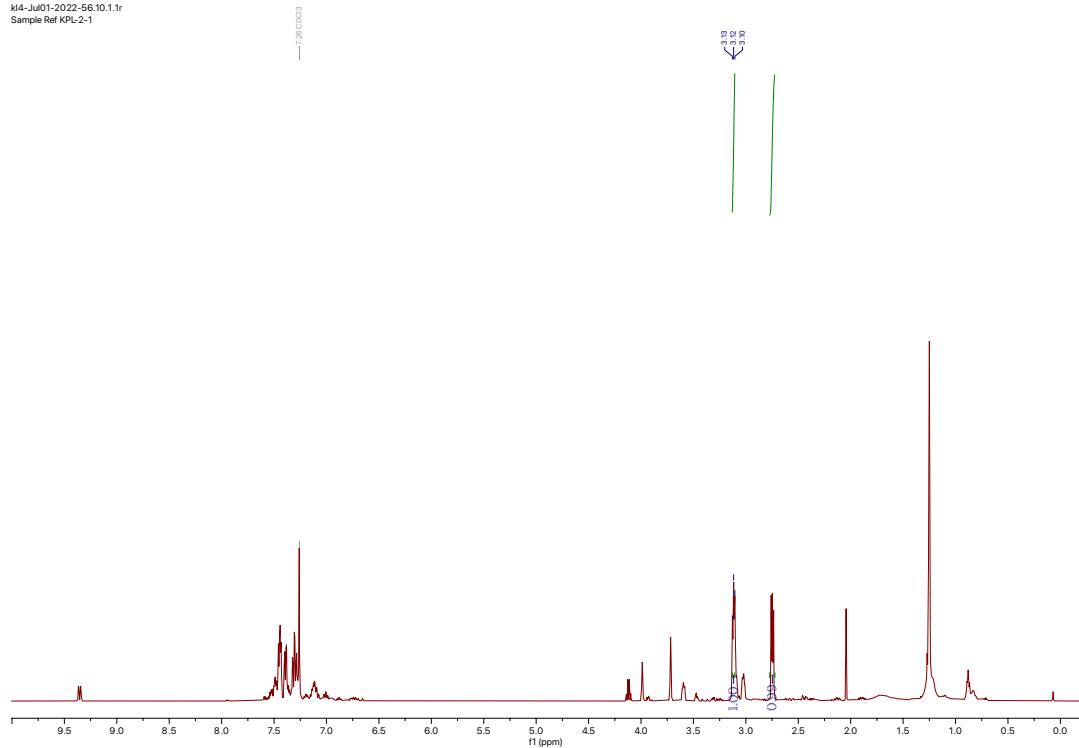
s1. ¹H NMR spectrum of 4-Fluoro-N-[(4-fluorophenyl)methyl]benzeneacetamide **6**



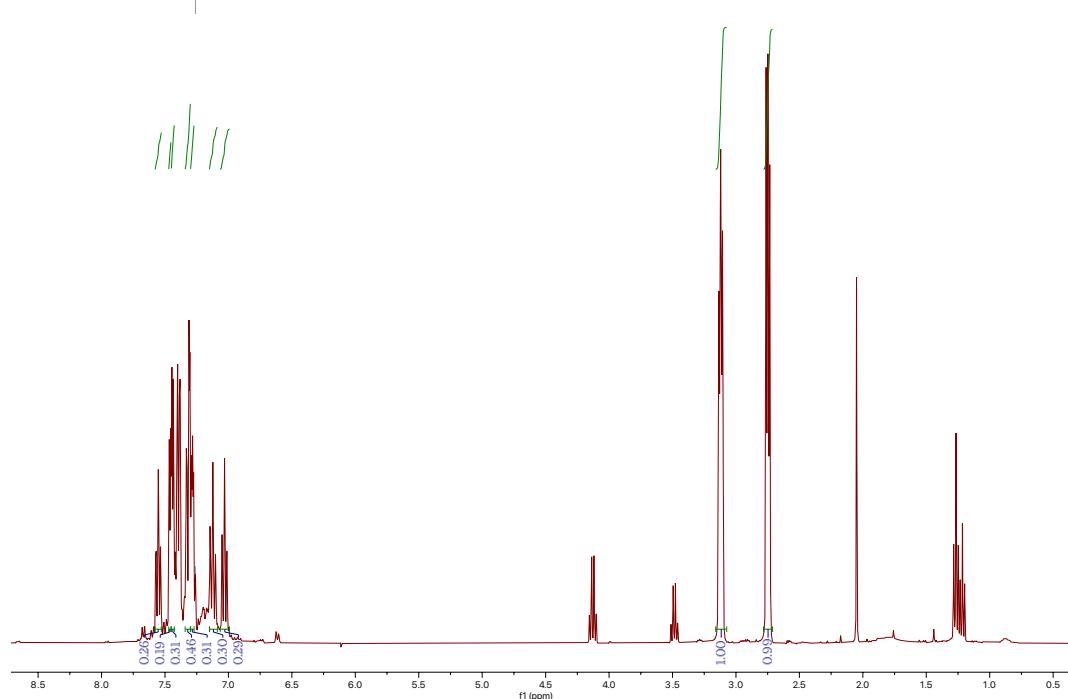
s2. ¹³C NMR spectrum of 4-Fluoro-N-[(4-fluorophenyl)methyl]benzenacetamide **6**



s3. FTIR spectrum of 4-Fluoro-N-[(4-fluorophenyl)methyl]benzenacetamide **6**

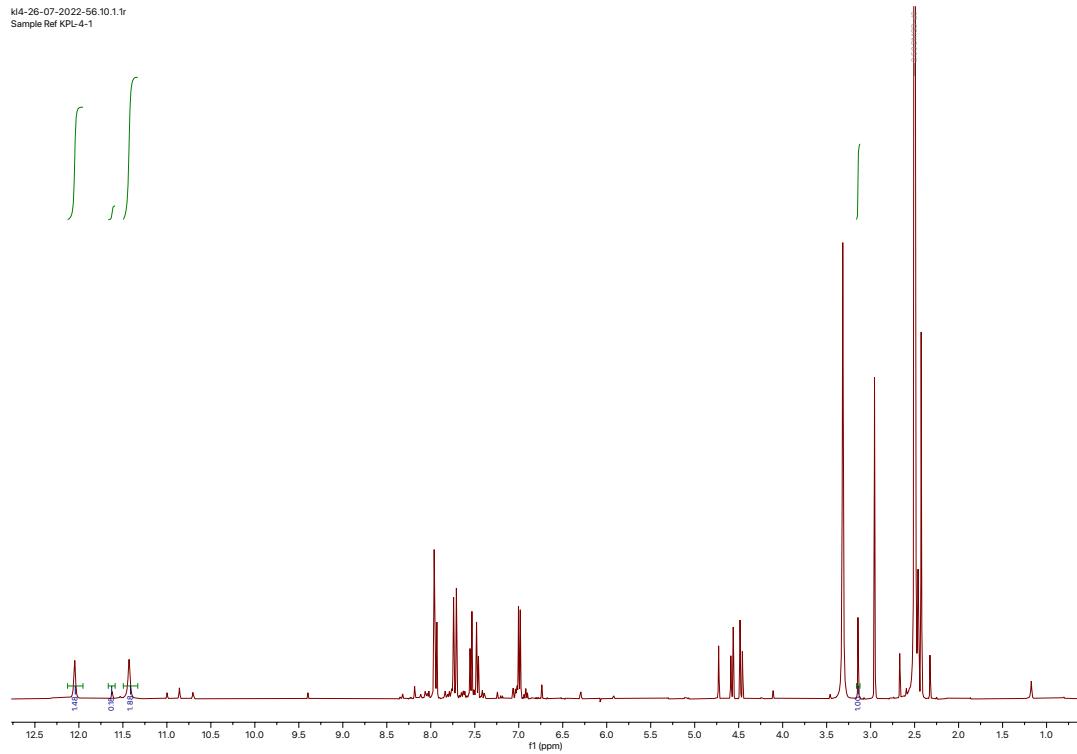


s4. ¹H NMR spectrum of post-column fraction in synthesis of 6-fluoro-1-(2-fluorophenyl)-1H-indene **14**



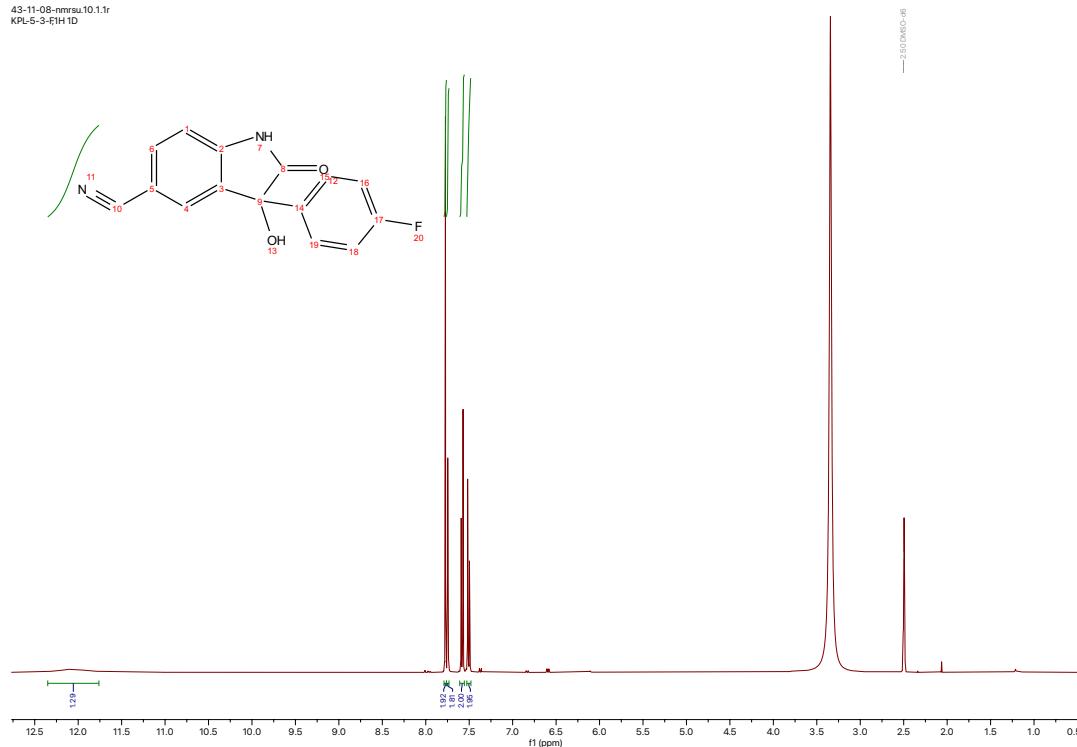
s5. ¹H NMR spectrum of crude in synthesis of 6-fluoro-1-(2-fluorophenyl)-1H-indene **14**

kl4-26-07-2022-56.10.1.1r
Sample Ref KPL-4-1

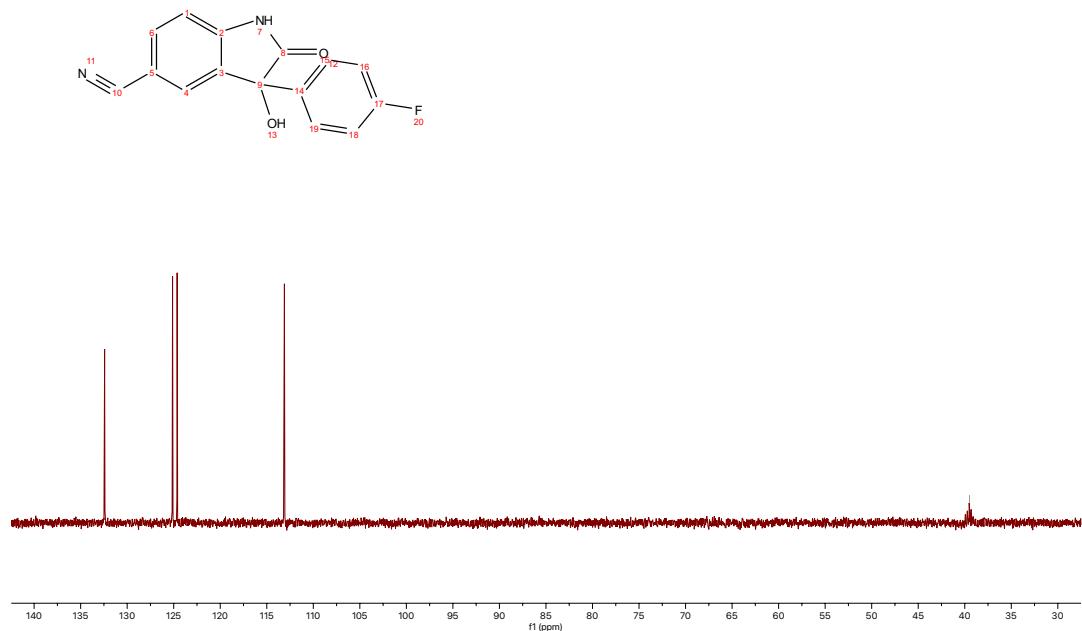


s6. ^1H NMR spectrum of crude in synthesis of 2,3-dioxoindoline-5-carbonitrile **16**

43-11-08-nmrsu.10.1.1r
KPL-5-3-F1H 1D



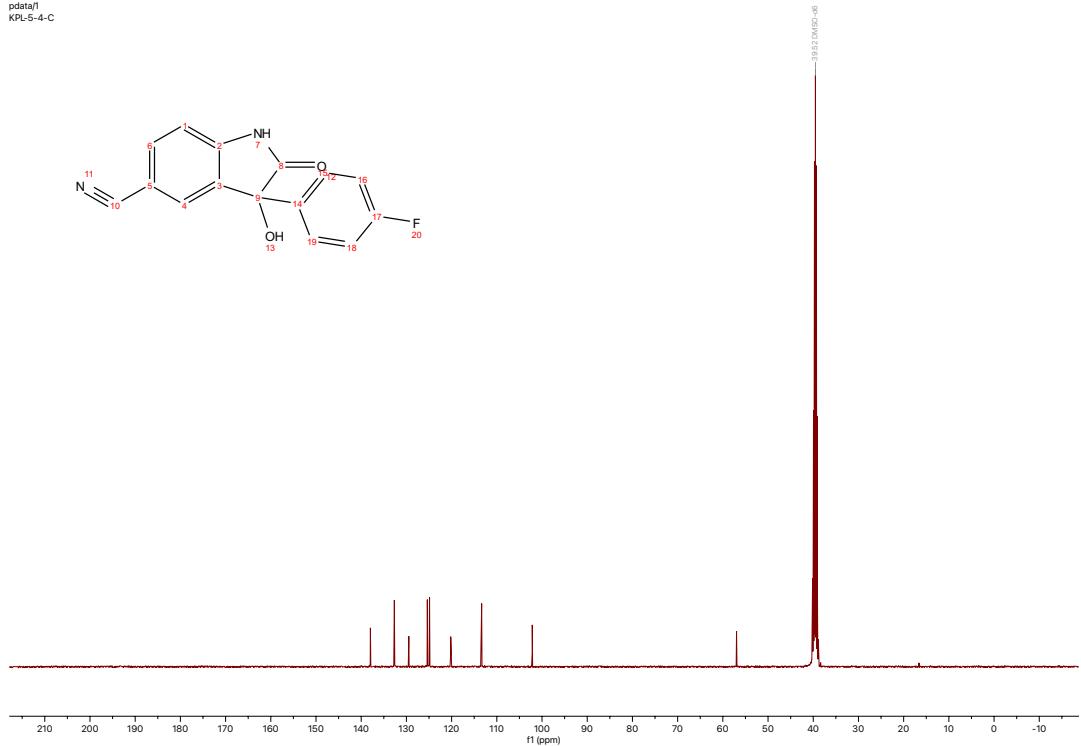
s7. ^1H NMR spectrum of crude in synthesis of of possible 3-(4-fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile **8**



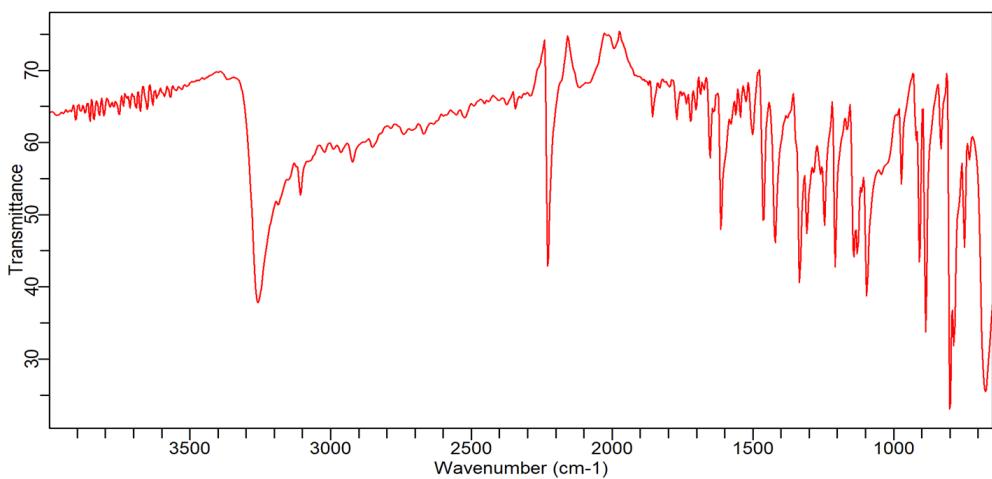
s8. Distortionless enhancement by polarization transfer-90 NMR Spectrum of possible 3-(4-fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile **8**



s9. ¹H-¹H Correlation spectroscopy NMR spectrum of possible 3-(4-fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile **8**



s10. ^{13}C NMR spectrum of possible 3-(4-fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile **8**



s11. FTIR spectrum of possible 3-(4-fluorophenyl)-3-hydroxy-2-oxoindoline-5-carbonitrile **8**