Structure-based Druggability Predictions of RNA Binding Sites and Design of RNA Ligands

Illimar Rekand

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2022

UNIVERSITY OF BERGEN

Structure-based Druggability Predictions of RNA Binding Sites and Design of RNA Ligands

Illimar Rekand

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 15.09.2022

© Copyright Illimar Rekand
The material in this publication is covered by the provisions of the Copyright Act.

Year: 2022
Title: Structure-based Druggability Predictions of RNA Binding Sites and Design of RNA Ligands

Name: Illimar Rekand
Print: Skipnes Kommunikasjon / University of Bergen

Tarkus tuleb tasapisi,
tuleb tasahilju,
tarkuselt ei tasu loota
kohe valmis vilju,
pole tarkus ladvaubin,
mida haarad käega,
ei saa tarkust pähe panna
võimu ega väega,
ei saa tarkust krati kombel
külast kokku kanda
ega selle tabamiseks
rasket raha anda.

Olivia Saar

Scientific environment

The work presented herein was carried out at the Department of Biomedicine, University of Bergen, Norway, under the supervision of Professor Ruth Brenk (Department of Biomedicine, University of Bergen, Norway) and at the Department of Chemistry, University of Bergen, Norway, under the supervision of Professor Bengt Erik Haug (Department of Chemistry, University of Bergen Norway).

The work has been financed by the Faculty of Medicine, University of Bergen and Norsk Forskningsråd (grant number 261961 and 294594) Funding was also received from The Norwegian Graduate School in Biocatalysis (BioCat), the Norwegian Chemical Society (NKS), Norwegian Ph.D School of Pharmacy (NFIF) and the Norwegian Biochemical Society (NBS) to participate in scientific meetings. Docking calculations related to DrugPred_RNA were performed on resources provided by UNINETT Sigma2. the National Infrastructure for High Performance Computing and Data Storage in Norway (Application NN9376K).

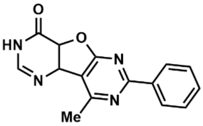
Abstract

The threat of the post-antibiotic era is looming and addressing this crisis requires access to new classes of antibiotics. Riboswitches are *cis*-gene regulating folded structures of non-coding RNA, which act through conformational changes induced after binding to small molecules. As riboswitches have been identified as a key component in bacterial biology and are essentially absent in mammalian cells, they constitute promising targets for novel classes of antibiotics

Herein, we have explored the possibility of addressing riboswitches as potential drug targets. A drug target needs to be relevant for a given disease and needs to be mofiable by either biological molecules or small molecules. In the latter case, the drug target needs to be able to bind small drug-like molecules with high affinity. Such binding sites are referred to as “druggable”. Previously in the Brenk group, a predictor (DrugPred) was developed to identify such binding sites in proteins. As part of this thesis, this predictor has been extended to be compatible with RNA binding sites. Due to the paucity of validated druggable RNA binding sites, the predictor, DrugPred_RNA, was trained on protein binding sites with descriptors that are applicable for both protein and RNA binding sites. It was found that this version was able to make predictions with high accuracy and precision, which were comparable to the last iteration of DrugPred. After applying this predictor on a large swathe of ligand-containing RNA crystal structures deposited in the Protein Data Bank, we have evaluated the predictions of DrugPred_RNA and found that known druggable RNA binding sites were classified correctly. The predictions of DrugPred_RNA were also found to be highly robust, and overall, binding sites which were found to be both similar in conformation and binding site composition were given the same prediction. However, it was also found that conformational changes could severely affect the outcomes. Further amon the RNA binding sites predicted to be druggable, many riboswitches were found, underlining this class of macromolecules as a potential target

for new antibiotics. Many of the druggable structures were also ribosomal binding sites, which could also guide the development of new antibiotics. Overall, DrugPred_RNA was found to be a useful tool for the prediction of druggable RNA binding sites

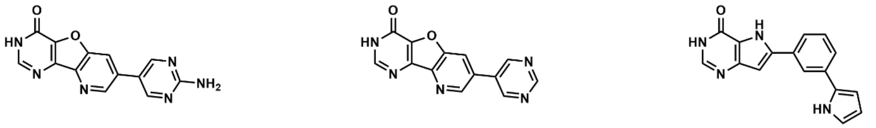
Among the riboswitches which were found to be druggable was the FMN riboswitch. This riboswitch regulates the expression of genes involved in biosynthesis of riboflavin and binds to flavin mononucleotide (FMN). The FMN riboswitch belongs to one of the most studied classes of riboswitches, and several potent binders have been identified. It has also been shown that these binders are able to inhibit growth of bacteria, and this, combined with its prevalence in human pathogens, including strains associated with nosocomial infections, makes it one of the most relevant targets to pursue. Using structural bioinformatics, we have designed compounds *in silico* and predicted their binding modes. The driving forces of the design of these candidates were to create synthetically feasible compounds which retain what we believe are key interactions observed or postulated between the FMN riboswitch binding site and FMN, ribocil A (a compound discovered by Merck) as well as for **RSL-0035** (**Figure 1**, a compound discovered in previous work within the Brenk lab), while simultaneously exploring new interactions.



RSL-0035

Figure 1: RSL-0035, a previously discovered hit in the Brenk lab

We designed four series of compounds based on these findings and performed synthesis of these compounds (two of which are described within this work) and **RSL-0035**, which proved successful and will serve as a basis for the synthesis of future compounds. Some of the steps within these syntheses were optimized and several discoveries were made which is going to shape the future of the project. Several of the designed compounds were found to be active, and a few were also found to have relatively high affinities in the single digit micromolar range. The structure-activity relationships of the compounds were established, and the findings have resulted in the proposal of new compounds which are currently pursued in the Haug group.



RSL-0026
$K_D = 2.2 \pm 0.7\ \mu M$
LE = 0.38 kcal$\cdot$mol$^{-1}\cdot$heavy atom^{-1}

RSL-0027
$K_D = 5.9 \pm 0.3\ \mu M$
LE = 0.36 kcal$\cdot$mol$^{-1}\cdot$heavy atom^{-1}

RSL-0034
$K_D = 8.3 \pm 2.9\ \mu M$
LE = 0.34 kcal$\cdot$mol$^{-1}\cdot$heavy atom^{-1}

Figure 2: Compounds discovered in this study with single digit micromolar affinity. LE = Ligand efficiency.

Acknowledgements

I would have not been able to writie this thesis without the support and guidance from my two supervisors. Thank you to Professor Ruth Brenk, who offered me to apply for a Ph.D. project without even meeting me and who later gave me the possibility to work in her group while we were applying for a Ph.D. grant together. Later she has shown support and guidance through my Ph.D. project, showing me how to perform science of high standards, showing me how to code and believing that I had the capabilities to finish if I only set my heart to it. Thank you to Professor Bengt Erik Haug, who supervised my first envoy into science with my B.Sc. project in 2013, and has guided me all the way now until finishing my Ph.D, nearly a decade later, who shares my perspective on what makes organic chemistry interesting both in and beyond the laboratories we work in. To both of you I want to express gratitude for supervising me in a joint effort and for motivating me through the last five years.

To everyone in the Brenk lab, you have proved to be great colleagues to work with and I want to express sincere gratitude. I want to thank Thomas Wehler and Raphael Klein, who showed great friendship and taught me many things early on in my Ph.D. period. Dr. Yves Nkizinkiko, who undertook the daunting task of developing the SPR protocol which eventually found use in this body of work and Dr. Charis Georgiou, who has done more for my wellbeing than he maybe realizes himself. To Dr. Vipul Panchal, who always incites discussion and curiosity in our group. To Vladyslav Yadrykhinsky, who I first got to know through our fencing club and has shown to not be only a dedicated fencer and colleague, but also a good friend. Special gratitude is directed to Dr. Fahimeh Khorsand, who in her relentless search for results, obtained the results from the SPR experiments which are presented in this thesis. Also thank you to Khanh, who has cared for the wellbeing of the group through taking care of business in the lab and the numerous cakes which you have baked over the years.

such a pleasurable work environment. Special thanks to Arne Raasakka, with whom I forged an early friendship. I also want to mention Kai, Erik, Helene, Marte, Svein, Kunwar, Fredrik, Juha K., Juha V., Laura, Pouda, Anastasia, Johanna and Sturla for making Biomedicine such a great place to work.

To all of my friends and colleagues in Department of Chemistry, I want to express sincere gratitude. Coming back here in the middle of a pandemic to work with synthesis felt like coming home. I want to thank everyone who has guided me in working in the Chemistry laboratory, with special emphasis to Dr. Muhammed Zeeshan, who has contributed immensely to the FMN riboswitch project, and to Dr. Åsmund Kaupang, who together with Zeeshan, became mentors for me in the 4030 laboratory. To my other colleagues in the Haug group, I want to extend gratitude: to Louise Bergsjø Sand, who has shared office with me for a long time and still enjoys my company, to Beate Halsvik, whose dedication and loyalty is inspiring to all, to Nina Henne, whose friendship has stretched far and who always wants to chat with me about chemistry or whatever else comes to mind. To Ludvik O. Espeland, who has shared my experience with working between the Haug and Brenk labs and also contributed with critical reading of this work. To my other colleagues in Chemistry, who makes this such a great place to work, thank you! I would like to thank Frida, Francesco, Eirin, Sara and also all the members of the lunch group who nouris my extroverted personality’s needs.

Thank you to the leader of the Department of Chemistry, Knut Børve, who allowed me to finish the necessary laboratory work I needed to finish my doctoral work.

Thank you to Dr. Bjarte Holmelid who has gone above and beyond in providing high resolution mass spectra and providing detailed methods for this thesis. Thank you to the NMR group, José Carlos Reyes Guerrero, Dr. Jarl Underhaug and especially Associate Professor Nils Åge Frøystein, who has provided insightful discussion on NMR shift predictions, and went out of his way to help me.

To my mother, Tiina and my late father, Ilmar, who has always shown love and supported and believed in me, who has shown me my roots and the path forwards to science. To the rest of my family, Carmen, Peeter, Markus and Jürgen, thank you.

To my Anne, who is always by my side and shows me unconditional affection and support. I love you.

To the Hognaland family, and especially Karin, who accepted me in and has shown me support.

To my friends, who bring me so much joy and support, thank you so much for giving me an opportunity to forget the stress of daily life. Thank you, Fredrik, who picks up the phone to ask how I am doing and is always there for me. Thank you, Joakim, Oda, Malin, Anders, Magnus, Leo, Anni, Endre, Britt-Elin and Camille!

To Hedvig Sultson, my good Estonian friend who shares the plight of academia with me.

I would also like to dedicate this to everyone who found hardship and insecurity in their Ph.D. program; you are not alone.

Contents

Scientific environment 3

Abstract 4

Acknowledgements 7

Contents 10

Abbreviations 13

1. Introduction 14

1.1 The pre-antibiotic era *14*

1.2 The inception of chemotherapy and antibiotic discovery *15*

1.3 The emergence of resistance and related problems *18*

1.4 Druggability *21*

1.5 Computational prediction of druggability *24*

1.6 RNA as a potential drug target *25*

1.7 Examples of riboswitches as potential drug targets for antibiotics *31*

1.8 The FMN riboswitch *32*

1.9 Binding mode of FMN and other ligands *34*

2. Aims 39

3. Results and discussion 40

3.1 DrugPred_RNA *40*

3.1.1 Construction of DrugPred_RNA 40
3.1.2 Initial evaluation of DrugPred_RNA 43
3.1.3 Prediction on RNA binding sites 46
3.1.4 Evaluation of predictions of DrugPred_RNA on RNA binding sites 50
3.1.5 Visual inspection and manual assignment of binding sites. 50
3.1.6 Relation between drug-like ligands and corresponding binding site prediction 51

3.1.7 Robustness of DrugPred_RNA 60
3.1.8 Finishing remarks 64

3.2 Design of New Ligands 66
3.2.1 Design and predicted binding modes of RSL-0029, RSL-0026 and RSL-0027 68
3.2.2 Design and predicted binding modes of RSL-0012, RSL-0007, RSL-0015 and RSL-0014 71
3.2.3 Design and predicted binding modes of RSL-0031, RSL-0032, RSL-0033 and RSL-0034 74
3.2.4 Design and predicted binding mode of RSL-0036 76
3.2.5 Final remarks regarding *in silico*-designed compounds 77

3.3 Synthesis of proposed FMN ligand candidates 80
3.3.1 Synthesis of hit compound RSL-0035 80
3.3.2 Synthesis of in silico designed compounds RSL-0031, RSL-0032, RSL-0033 and RSL-0034 86
3.3.3 Synthesis of RSL-0036 97

3.4 Biological Data 101

4. Conclusions and further work 109

5. Methods and general experimental 113

5.1 DrugPred_RNA 113

5.2 Molecular Minimization and Ligand Design 120

5.3 Docking 120

5.4 General Synthetic Experimental Methods 121

5.5 Activity assay 123

5.6 Other 123
5.6.1 Molecular descriptors of *in silico* designed compounds in Table 7 123
5.6.2 HRMS analyses 123
5.6.3 NMR predictions in Table 10 125
5.6.4 Synthetic experiments 126
5.6.5 Graphics software 126

6. Experimental 127

6.1 *Experiments leading up to and including* ***RSL-0035*** *..... 127*

6.2 *Experiments leading up to and including* ***RSL-0031, RSL-0032, RSL-0033 and RSL-0034*** *..... 151*

__6.3__ Experiments leading up to and including __RSL-0036__ *189*

7. Appendix 212

8. References 221

Abbreviations

Most abbreviations used within the text of this thesis conform to the standard abbreviations and acronyms proposed by the *Journal of Medicinal Chemistry.*[1]

Below are listed some selected abbreviations which do not conform to the selection mentioned above.

A	Adenine /Adenosine
C	Cytosine / Cytosine
U	Uracil/Uridine
G	Guanine/Guanosine
FMN	Flavin mononucleotide
TPP	Thiamine pyrophosphate
SAM	*S*-adenosyl methionine
TAR	Trans-activation response
ML	Machine learning
SHAP	Shapley additive explanations
xgboost	eXtreme Gradient Boosting
DLID	Druglike density
QED	Quantitative estimate of druglikeness

1. Introduction

1.1 The pre-antibiotic era

Arguably the greatest achievement humanity has gained through modern medicine is the prevention of countless deaths by infectious diseases. While it is near impossible to produce an exact number of how many lives have been saved, the Centers for Disease Control and Prevention provide a crude death rate for infectious diseases in the United States over the 20th century.[2] At the start of the century, the number of deaths caused by infectious diseases and the number of all other death causes were almost equal, but today the former number has been reduced over twentyfold, leading to the almost complete annulment of infectious disease-related deaths in Western society, as can be seen in **Figure 3**. What is the cause for this severe reduction? A certain cause lies in a greater understanding of the underlying cause of infection. The leading theory for how diseases transferred between host patients up until the late 19th century was miasma theory, which stated that diseases relayed through foul air, not through bacteria or viruses.[3] While Leeuwenhoek first described bacteria as early as in 1676, it would take

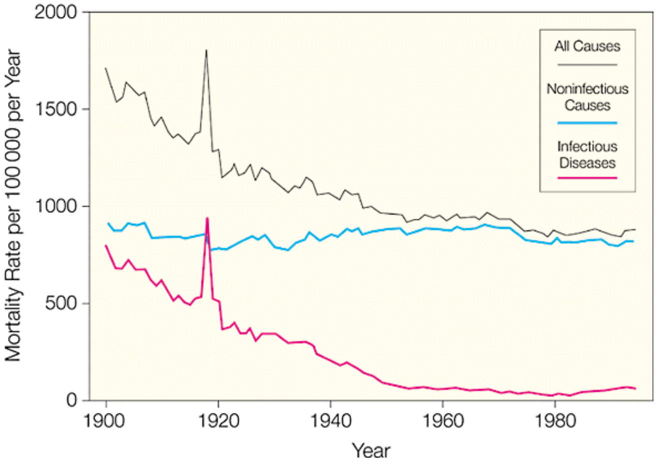


Figure 3: Total number of deaths in the United States in the 20th century. Figure taken with permission from Armstrong *et al.* [1]

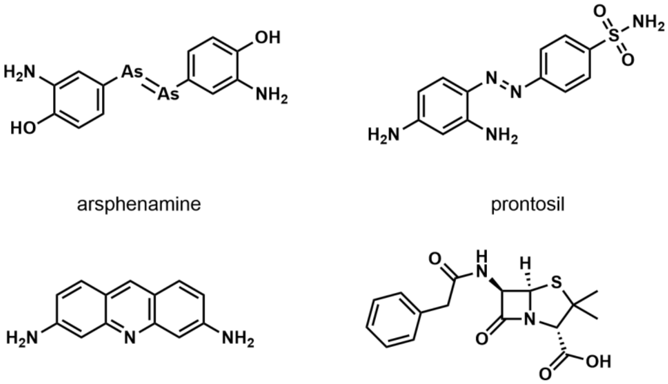
centuries before the connection was made between the proliferation of pathological organisms in host patients and outbreak of disease with the proposal of "germ theory" as put forth by Lister, Pasteur and Semmelweis in the late 1800s.[4–6] This new understanding led to idea that such disease was preventable through measures such as proper hygiene, sterilization and vaccination.

Due to improved general and personal hygiene, the number of infectious-related deaths declined before the introduction of effective antibiotics in the 1940s and 50s. However, these measures were not sufficient to prevent every of bacterial disease. Prior to the antibiotic age, patients suffering from bacterial infections, from for example *Streptococcus pneumoniae*, had a low chance of survival, and diseases such as tuberculosis had a mortality rate of 50%.[7,8]

1.2 The inception of chemotherapy and antibiotic discovery

It is likely human beings have unknowingly used bacteriostatic agents to treat disease for millennia. Records show that topical applications of moldy bread was used in ancient civilizations, and skeletal remains in Sudanese Nubia from A.D. 350-550 were found to have trace amounts of tetracycline.[9,10] While the first known use of a chemical to prevent a bacterial disease can be attributed to Lister's use of carbolic acid (phenol) as an antiseptic, it was not until the beginning of the 20th century the idea appeared that a chemical agent could actively cure a disease. As more knowledge of the bacterial life cycle increased, it was speculated that a specific substance could be able interfere with it, which laid the foundation for what is known today as chemotherapy.

In many ways, the birth of modern medicine started with the inception of modern antibiotics. Arsphenamine (**Figure 4**), an arsenic-containing drug developed in 1910, was the first true synthetic drug, *ie*. a compound intended for medical use which was not naturally occurring and proved effective for the treatment of syphilis.[11,12] With proflavine and prontosil, the first broad-spectrum antibiotics were introduced in the 1930s.[13] The latter agent was the first of the sulfa drug class, which is still in use today. The sulfa drug class is characterized by its benzene ring with a sulfonamide and amine moiety (sulfanilamide), which was later found out to be metabolized from a diimide group. This marks the first discovery of a prodrug.



arsphenamine
prontosil
proflavine
penicillin G

Figure 4: Structures of some early antibiotics

These early findings now marked the start of the antibiotic age, which has led to a series of discoveries (**Figure 5**) The discovery of penicillin isolated from *Penicillium rubens* molds had a huge impact on the field of medicine for two reasons: 1) the compound displayed extraordinary potency while showing little to no toxicity 2) The idea that nature already provided potent antibacterial agents kickstarted a screening campaign to discover novel antibiotics from microbial cultures around the globe.[14] While penicillin was discovered as early as 1928, it did not find widespread use until much later, as it was unstable and production proved to be difficult to upscale.[15] The eventual structural

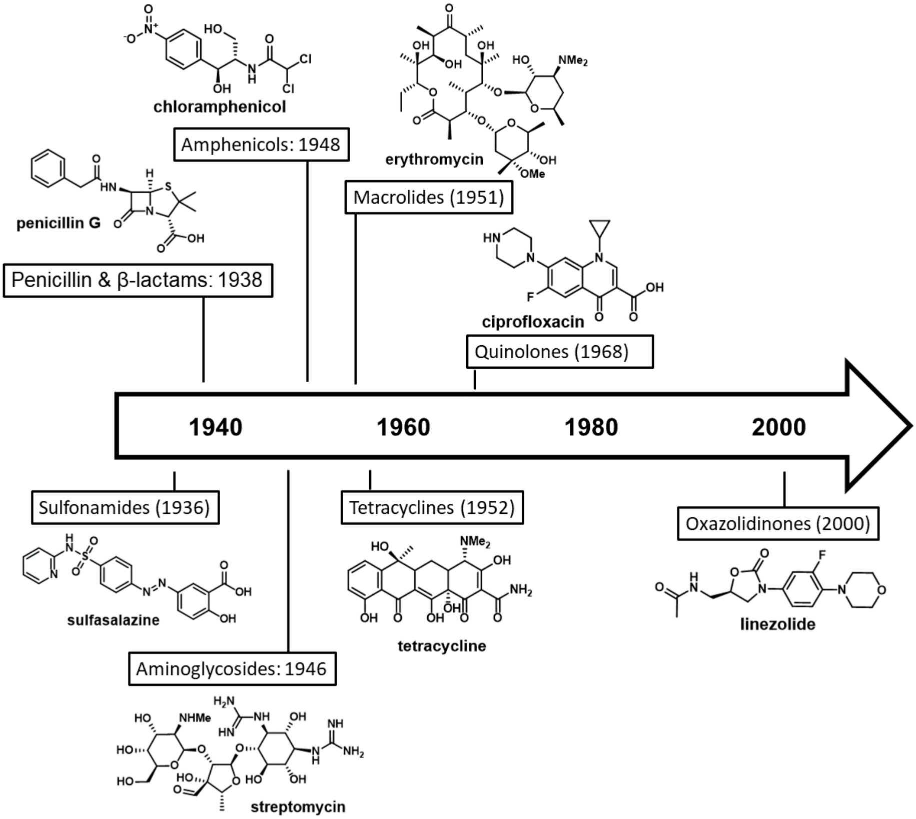

chloramphenicol
Amphenicols: 1948
erythromycin
Macrolides (1951)
penicillin G
Penicillin & β-lactams: 1938
ciprofloxacin
Quinolones (1968)
1940
1960
1980
2000
Sulfonamides (1936)
sulfasalazine
Aminoglycosides: 1946
streptomycin
Tetracyclines (1952)
tetracycline
Oxazolidinones (2000)
linezolide

Figure 5: Timeline displaying discoveries of selected antibiotic classes shown with examples of each respective class.[12]

elucidation and comprehension of its chemical properties, development of a semi-synthetic route and lyophilization technology resulted in that the β-lactams becoming one of the most ubiquitous antibiotics of modern age.[16,17] However, a series of hurdles were still in place despite the great potential of this new discovery: the production of penicillin at the time was still cumbersome, the drug still lacked efficacy against certain infections, and reports of allergies raised the need for other classes of antibiotics.[18] Streptomycin, isolated from *Streptomyces griseus* in 1945, was the first discovered aminoglycoside that was also the first antibiotic effective in treating tuberculosis. The following discoveries in the period of 1950 to 1960 is described as the golden age of antibiotics, as half of all known antibiotics still used today were first described in this time.[17,19]

As penicillin and other β-lactam antibiotics inhibits the cross-linking of the peptidoglycan wall in bacteria, the newly emergent classes in the successive decades targeted completely different aspects of the bacterial lifecycle. Several classes, such as the amphenicols,[20] macrolides[21], aminoglycosides[22] and the tetracyclines[23], inhibit protein synthesis by targeting different sites of the ribosome, vancomycin obstructs the construction of the bacterial peptidoglycan cell wall,[24] rifampicin inhibits RNA synthesis by targeting the bacterial polymerase,[25] while the quinolones target bacterial DNA replication.[26] Thus, the antibiotics we have today are as varied in action as in their chemical structures.

1.3 The emergence of resistance and related problems

The advent of antibiotics was hailed and was described as a “magic bullet” of medicine, which could now be easily used to remedy diseases previously thought of as a death sentence. Thus, they found widespread use. The high need for these drugs in modern society has resulted in for example multi-metric tonnage annual production of 6-aminopenicillenic acid, the precursor to most penicillins.[27,28] However, as the use of antibiotics increased, so did also the reports of antibiotic resistant strains. For instance, from the 1970s to the 1990s, less than 100 β-lactamase enzymes were identified, but by 2010 this number had skyrocketed to a near 1000.[29] Around 19.000 deaths were caused in 2005 in the US by methicillin-resistant *Staphylococcus aureus* (MRSA), which is also associated with an annual healthcare cost of around $3 billion.[19,30] In a broader picture, the total annual excess healthcare cost related to resistant infections in the United States and European Union is estimated to be $20 billion and 1.6€ billion, respectively.[19] A study showed the total numbers of antibiotic resistance-related infections in Europe in 2015 to be >670.000, with >33.000 attributable deaths.[31] A statistical model estimated that the global death tally associated with antimicrobial resistance (AMR) was 4.95 million in 2019.[32]

The rise in antibiotic resistance has been attributed to several factors: 1) The overuse of antibiotics in clinical applications, for example in treatment of non-bacterial related diseases, or overly long or short regimens of administration or improper treatment.[19,33] 2) The use of antibiotics for the intent of prophylaxis, metaphylaxis and growth promotion in the food industry, which account for up to 50% of all antibiotic use, has also been a key factor in the drastic emergence of antibiotic resistance.[34,35] 3) The dearth of advances in novel antibiotic classes has made the situation even more severe with the arrival of class-wide resistance. The lack of interest from the pharmaceutical industry to develop new antibiotics is explained by the fact that antibiotics are only prescribed for limited durations, making them less profitable, combined with the fact that newly developed antibiotics are typically held in reserve and only used when established antibiotics fail to work.[19] After factoring in the gargantuan cost of discovery and development of new drugs, antibiotics stand as a less desired area for the pharmaceutical industry. 4) Although the human contribution to the problem is significant, the occurrence of resistance can appear without human intervention. Genes associated with resistance (such as β-lactamases) are found to transfer horizontally between microbial communities as well as vertically, and phylogenetic analyses suggests the emergence of cross-transfer of such resistance-related genes dates back to long before the discovery of penicillin and other antibiotics.[19,36–38]

The reported numbers of antibiotic resistance have increased drastically over the last few decades. The phenomena of resistance development are far from unknown, as the first cases of penicillin resistance in a clinical setting were already observed as early as in 1942. The effects of unrestrained use of penicillin were also warned by Sir Alexander Fleming in his Nobel lecture in 1945:

> *"*It is not difficult to make microbes resistant to penicillin in the laboratory by exposing them to concentrations not sufficient to kill them, and the same thing has occasionally happened in the body. The time may come when penicillin can be bought by anyone in the shops. Then there is the danger that the ignorant man may easily underdose himself and by exposing his microbes to non-lethal quantities of the drug make them resistant.*"*[39]

Three classes of resistant pathogens have emerged as the key major threats to public health: 1) MRSA, 2) several multi-drug resistant (MDR) and pan-drug resistant (PDR), and often nosocomial, Gram-negative bacteria, specifically *Acinetobacter baumannii, Escherichia coli, Klebsiella pneumonia* and *Pseudomonas aeruginosa* and 3) MDR and extremely drug-resistant (XDR) *M. tuberculosis* strains.[19,29,40,41] The World Health Organization (WHO) approved a global action plan in 2015 to combat the problem, calling it "a crisis that must be managed with the utmost urgency", and later also compiled a list of bacteria for which new antibiotics are most urgently needed.[42,43] Furthermore, it is estimated that antimicrobial resistance may lead to nearly 10 million deaths globally per year by 2050, surpassing cancer, the leading cause of death today.[44] As of today, there is a special need for novel targets for antibiotic activity, as since the 1960s, only six new classes of antibiotics have been introduced, and none of which have activity against Gram-negative bacteria.[19] The "l0 x '20 initiative" by the Infectious Diseases Society of America in 2010 sought out to combat this lack of innovation, whereupon the development of 10 novel and effective antibiotics before 2020 was called out for.[45] A final report on the initiative from 2019 stated while a great number of new antibiotics were in clinical trials, a majority of approved agents were modifications of previously discovered classes and a minority were targeting Gram-negative bacteria.[46]

1.4 Druggability

With antibiotic resistance on the prowl, the need for new antibiotics with novel mechanisms of action rises. A major issue the pharmaceutical industry is facing with development of new drugs is the associated risk and cost from inception to release on the market. A >90% overall estimated failure risk is found in drug development, with the highest rates found for previously undrugged targets.[47–49] Moreover, a large number of these failures have been attributed to the fact that the project's relevant target simply failed to meet the necessary physicochemical criteria associated with high-affinity binding or inability to bind druglike molecules.[50] Factoring in the resource demands for development to 12 years with an estimated cost of $1150 billion U.S per drug. underlines the importance of eliminating likely failures early in the drug discovery process.[51] Consequently, the development of a predictor able to discriminate binding sites with favorable physicochemical properties from unfavorable is imperative to avoid a massive sunken cost into a failed project.

Druggability refers to a target's likelihood to bind a small, organic, orally available drug-like molecule with high affinity.[52,53] Originally coined by Hopkins *et al.*, "the druggable genome" referred to proteins with corresponding genetic sequences with close similarity to already known drug targets, which are able to bind small molecules which are "rule of 5"-compliant.[54] Other synonyms have been used, such as "ligandability", "bindability" and "(chemical) tractability", but the term "druggability" is used in this work as it is the most prevalent term used in the literature.[52,53]

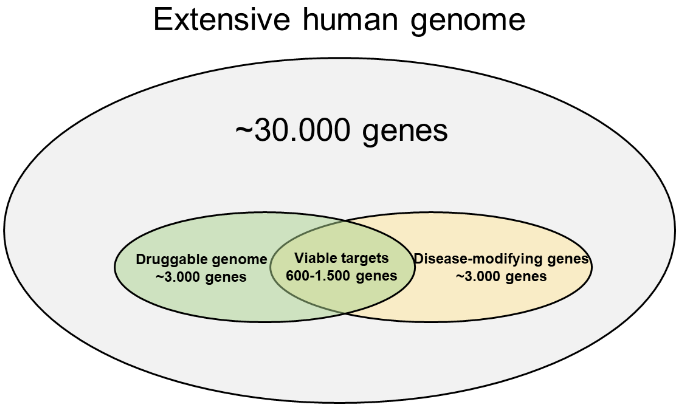


Figure 6: The original analysis by Hopkins found that only <5% of the human genome contains viable druggable protein targets.[93,94]

In the original analysis by Hopkins *et al.* which coined the druggability term, less than 5% of the human genome (**Figure 6)** was found to be associated with druggable proteins.[54,55] The fact that a target not only has to fulfill druggable potential, but also must be associated with a disease, underlines how narrow the biological space of viable druggable targets is. The main challenge associated with this topic is the discrimination between what is druggable and what is not. The main plights of defining druggability lies within the abstract nature of its concept, as it is not an absolute property and the paucity of data about what defines the less druggable space.[53] While originally druggability was characterized through corresponding genome sequences, this approximation has been replaced by a more structure-focused approach. This shift came with the realization that genomic analysis fails to account for alternative splicing and post-translational alterations, and the fact that some targets could be missed if sufficient genomic information is missing. Thus, the focus was shifted away from a class-wide definition of druggability and on the assessment of individual members.[56]

Previous attempts at defining druggability has been through the use of high-throughput screening (**HTS**) of different protein targets to see which yield multiple, structurally diverse hits.[57] Both GlaxoSmithKline (**GSK**) and Astra Zeneca (**AZ**) have performed extensive screening campaigns on multiple protein targets grouped together by classes explicated by function (ligases, kinase-1, kinase-2, transferases, etc.), and defined druggability by the relative hit rate within a class.[58,59] GSK's approach to defining hits were only to include genuine validated hits, while AZ defined hits as any compounds providing a signal above a certain response level. The question then arose: should druggability be defined as a protein's ability to bind any small molecule or only true validated hits? Another obvious drawback of this method, apart from being extremely costly, laborious and resource demanding, is the limited scope of chemical space a chemical library can scan. While a typical HTS library can contain thousands or even millions of unique compounds, this would not span more than a 10^{-16} % of the synthetically accessible chemical space, which is estimated to be composed of up to 10^{24} unique compounds.[60] Therefore, the move towards smaller libraries with more structurally diverse compounds was heralded. These fragment libraries, which typically count compounds in the hundreds with a molecular weight below 300 Da, span chemical space more effectively, with the added benefit of providing higher hit rates, as the threshold affinity for a "hit" is lower than for a traditional HTS campaign.[61] AZ and other companies have demonstrated a correlation between poor hit rates in fragment screens and difficulty in obtaining high-affinity ligands.[52,62,63] One of the main challenges this methodology faces, is the low affinity usually obtained by fragment hits necessitates methods with high sensitivity to differentiate false positive hits from actual hits. From a drug-discovery perspective, performing a fragment screen, even however small the library is, on a multitude of targets to assess their druggabilities quickly becomes a daunting task. LifeArc has presented an automated workflow which improves the efficiency of a screen through the use bioinformatic triage to eliminate poor candidates and small scale high-throughput protein production enabled one target to be screened per day.[64]

1.5 Computational prediction of druggability

The availability of highly detailed three-dimensional structures of proteins (and other biological macromolecules) has opened the possibility for computational prediction of the druggability of binding sites. This is enabled through the calculations of physicochemical descriptors of binding sites. The common approach in computational assessment of druggability has been 1) identifying the binding sites, which typically has been conducted through a cavity detecting software, 2) describing the properties of the cavity and 3) comparing them to known data to predict the druggability of a given pocket.[57,65–68] Common binding site descriptors are volume size, curvature and other geometric attributes, both relative and total polarity and lipophilicity, surface area, enclosure, and many more.[69] A predictor can then be trained on the basis of these data.

The advent of machine learning (ML) methods has opened the possibility to use larger swathes of data to more accurately describe which molecular properties portray the differences between druggable and less druggable sites. Several approaches using ML methods to predict druggability have been made, based on both sequence and structural data.[70] The earliest structure-based prediction method was developed by Hajduk *et al.* in 2006, where training data was derived from a heteronuclear NMR-based fragment screening data on 38 pockets from 23 different proteins to categorize the pockets. After pockets had been identified using a flood-fill algorithm, their parameters were calculated and used for a regression analysis and predicted on a test set of 72 proteins.[63] While the method served as a great proof of concept, accuracy (58%), recall (93% and 52% for druggable pockets and less druggable pockets, respectively) and precision (66%, 88%) were still unsatisfactory. Schmidtke *et al.* created a druggability predictor using a bootstrap ML method based on a nonredundant druggability data set, which offered an improved accuracy of 70% (recall: 88%, 44%, precision: 66%, 78%).[57] Druglike-density (DLID) which explores only three pocket variables (volume, buriedness and hydrophobicity) to provide an index score, has also been developed.

While the receiver operating characteristic (ROC) area under curve provides a reasonably high accuracy (0.90), the method does not clearly discriminate between druggable and less druggable sites and categorizes binding sites in between as “difficult”.

Previously in the Brenk group, DrugPred has been developed as a druggability predictor of protein binding sites.[71,72] Description of physicochemical properties of binding sites from X-ray crystal structure are assessed through creation of a negative mold of the binding site, which is created by docking of a plethora of approved drug molecules in a binding site after removal of the cognate ligand, and combining the atoms of ligand poses which scored above a certain energy threshold into one combine molecule. This molecule, hereafter called “superligand”, can then be used to calculate various descriptors such as volume, buriedness, polarity, *etc*. The predictive performance of DrugPred was found to outperform previously reported druggability methods, with an accuracy of 91% (recall: 96%, 83%, precision: 90%, 93%). Another novelty with DrugPred is the data set the predictor was derived from, which is a collection of nonredundant druggable and less druggable protein binding sites (NRDLD) to avoid an overfitting bias in the final prediction method. When DrugPred was applied as a prediction method on previously reported dataset (*e.g.* the Hajduk dataset), improved prediction performance were observed. As such, DrugPred was found to be a more reliable druggability predictor than what had been previously created.

1.6 RNA as a potential drug target

With the impending threat of bacterial resistance, the search for new targets for antibiotics continues. Traditionally, protein binding sites have been the classic target for developing drugs.[73,74] Nucleic acids have remained a lesser explored domain for drug development, but there are still numerous examples of FDA approved drugs which targets these. Most of these act on DNA, either through alkylation (such as cyclophosphamide[75]), complexation (cisplatin) or intercalation (doxorubicin[76]). A number of RNA-targeting drugs act on the ribosome (and are found to bind more often to ribosomal RNA rather than ribosomal protein)[77,78], such as tetracyclines[23], aminoglycosides[22] and macrolides[21]. One of the more recent additions to the antibiotic classes is the oxazolidinone class, with linezolid as its first-in-class, which also inhibits initiation of protein synthesis by targeting the bacterial ribosome[79–82] RNA's role as a promising drug target has evolved in tandem with our growing knowledge of the roles and properties of RNA.[83] This includes the ability to form three dimensional folded structures and ability to selectively recognize small molecules[84–86]. Warner *et al.* has written an extensive review, arguing that RNA molecules which are specifically folded and able to bind highly drug-like compounds are the most promising targets to pursue for further investigation.[84] A druggability assessment of several PDB-deposited RNA X-ray crystal structures using DLID has been published, which concluded that many RNA binding sites are suitable for ligand development.[87]

RNA is a polymeric structure composed of nucleotide monomers and with a backbone composed of ribose sugar units (**Figure 7A)** interlinked with phosphodiester groups which are elongated in the 3'→5' direction. The sidechains extend from the 1' position on the sugar moiety and four different heterocyclic nitrogen-enriched nucleobases constitute the differences between RNA nucleotidic residues: guanine (**G**), adenine (**A**) and cytosine (**C**) and uracil (**U**) (**Figure 7B)**. The nucleobases are paired through specific and complementary Watson-Crick base-pairing mediated by hydrogen bonding between the endo- and exocyclic nitrogen and carbonyl oxygen atoms in the heterocyclic structures **(Figure 8A**). Thermodynamic stability is further enhanced through hydrophobic interactions through π-π stacking of the nucleic acids **(Figure 8B**). Additionally, the 2'-hydroxyl group provides further support with the ability to provide further hydrogen bonding. The sum of all of these interactions allows RNA to fold and form stable 3D tertiary structures **(Figure 8C**).

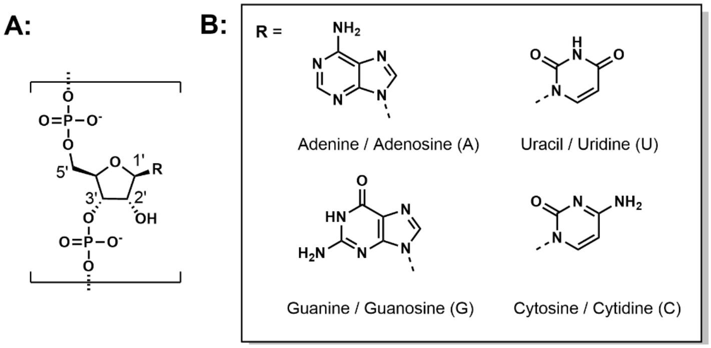


Figure 7: A) The ribose backbone of RNA, where the various nucleobases are attached on the 1’ position. Each nucleotide residue is interconnected with phosphodiester groups in the 3- and 5- position **B)** Nucleobase name / Nucleotide name (One-letter code)

RNA strands vary greatly in length, from small interfering RNA in *D. melanogaster* (21-24 nucleotides)[85] to the average length of mRNA transcripts (~1,250 nucleotides in *S. cerevisiae*[89]), but when compared to DNA is generally much shorter (for example, the human chromosome is one single DNA molecule composed of 247 million continuous nucleotides).[90]

RNA is divided into three subclasses; (1) messenger RNA (**mRNA**), which consists of shorter to intermediately long strands of encoded RNA, and act as intermediaries of genetic information between DNA and the ribosomal factory[91], and are believed to have no strong secondary structure features, (2) transfer RNA, (**tRNA**) which assist in the process of translation and expression of mRNA in the ribosome through specific folding and binding to protein structures (**Figure 8C**) and (3) ribosomal RNA (**rRNA**), which are longer chains of RNA, which serves both a structural purpose in the ribosome and is also enzymatically active in the formation of peptide bonds in the translational processing of mRNA.[92] Although a majority of RNA conforms to the mentioned subclasses above, it also has been adapted to other roles. Ribozymes are enzymatically active RNA molecules which are specifically folded and can catalyze certain biochemical transformations. One example is the hammerhead ribozyme, which catalyzes the cleavage of phosphodiester in the backbone of RNA.[93] Small interfering RNA (siRNA) are short, double-stranded RNA moieties which interfere with expression of certain complementary regions of DNA.[94]

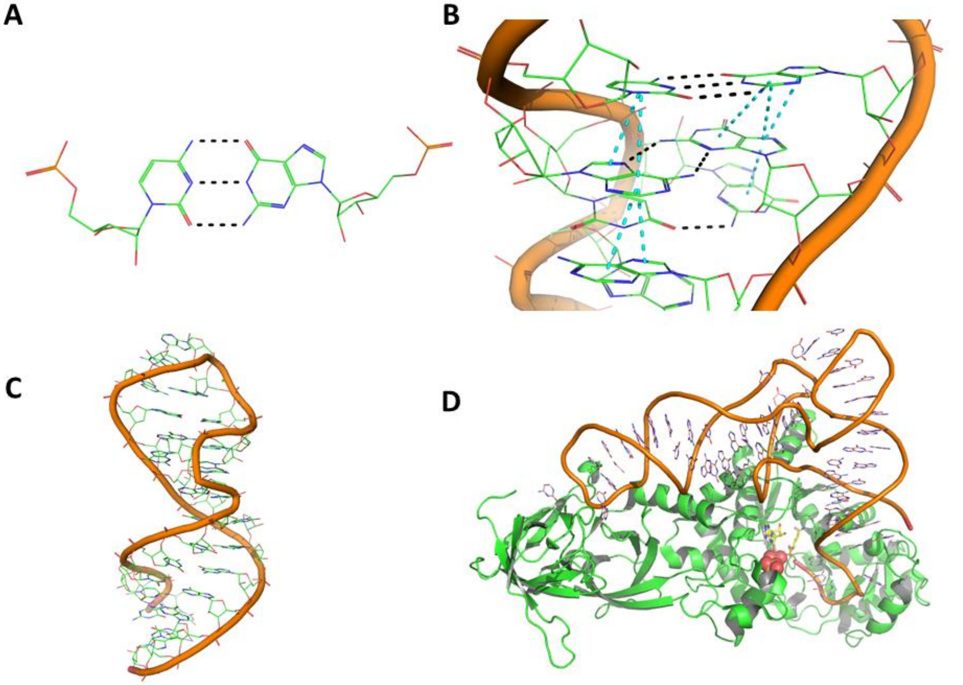


Figure 8: A) Close-up of Watson-Crick base-pairing between a guanidine and cytosine residue. Residues shown as thin sticks. Hydrogen bonding shown as black dashes (PDB ID: 6y0y) **B)** Key interactions in an RNA structure (PDB ID: 6y0y). Backbone shown as orange tubes, residues as thin sticks, hydrogen bonds shown as black dots and pi-stacking shown as cyan dashes **C)** Zoomed out perspective of whole RNA loop structure **D)** single-stranded tRNA bound to protein (PDB ID: 1zjw). Nucleic acid backbone shown as light green sticks, hydrogen bonding shown as black dashes, protein shown as green cartoon.

In recent years another type of non-coding RNA entities has become of interest, which are riboswitches. These entities, which are found in the 5’ untranslated region of mRNA, are able to regulate gene expression by specific binding to a cognate ligand. They appear almost exclusively in bacteria.[95–97] Riboswitches are folded into specific three-dimensional structures, and are divided into two subdomains; an aptamer region which forms binding sites which can selectively bind ligands ranging from single atoms to small molecule metabolites (**Figure 9**), and an expression platform which regulates expression of associated genes upon the binding of a ligand. Multiple classes of riboswitches have been discovered and are as such classified primarily by their natural ligands and secondarily by the fold of the RNA aptamer. The cognate ligands include the amino acid lysine, the enzyme cofactor thiamine pyrophosphate (**TPP**), guanine, the enzyme cosubstrate *S*-adenosyl methionine (**SAM**), and phosphorylated vitamin B_2, also known as flavine mononucleotide (**FMN**) among others. The purpose of these RNA structures is gene regulation, where binding to a small molecule causes a conformational change, which enables an adjacent expression platform to either become available (“ON”-riboswitches) or unavailable (“OFF”-riboswitches) for either transcription or translation.[98–100] Normally, this regulation takes place in a feedback loop mechanism, where the synthesis of a metabolite (such as TPP) results in an increased concentration of the cognate ligand, which then binds to the riboswitch and at a threshold concentration level shuts down expression of metabolite producing genes. Riboswitches are also extremely specific. For example, the TPP riboswitch is able to discriminate between its cognate ligand and thiamine monophosphate with a 10/1000-fold greater activity.[101]

Over 28 riboswitch classes have been experimentally validated, with some riboswitches found in more than 5000 bacterial species, including human pathogens, and riboswitches are estimated to regulate up to around 2.2% of the bacterial genome (*B. subtilis)*.[102–104] Their ability to selectively bind small molecules with high affinity combined with their widespread abundance in bacteria makes them attractive targets for antibiotics.[83,95,96,99,105]

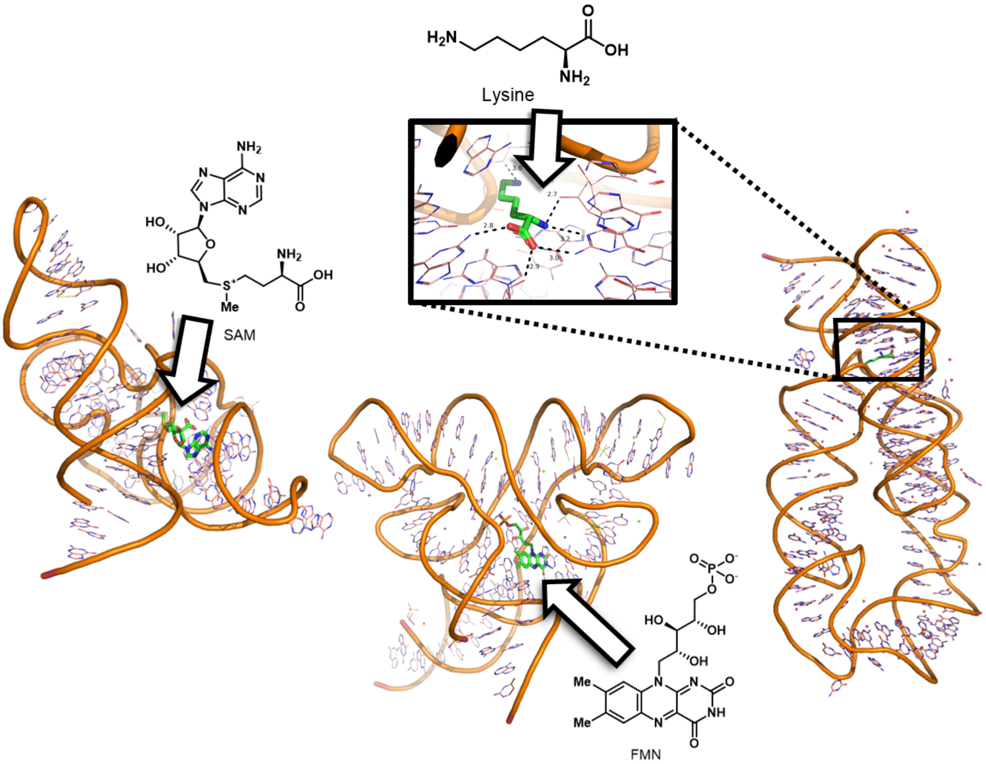


Figure 9: Examples of three riboswitches with their natural ligands. From left to right: the SAM riboswitch (PDB ID: 3gx5), the FMN riboswitch (2yie) and the lysine riboswitch (PDB ID: 3d0u)

1.7 Examples of riboswitches as potential drug targets for antibiotics

Binding to riboswitches has proven to be successful in killing bacteria in several cases. One of the earliest examples is pyrithiamine, which was discovered to be toxic to both fungi and bacteria before the identification of the first riboswitch.[106] This compound, which was originally prepared as a structural analogue of TPP, is intracellularly phosphorylated to give the compound pyrithiamine pyrophosphate (PTPP), which has a similar binding affinity to the TPP riboswitch as TPP (160 nM *vs.* 50 nM). Bacteria resistant to pyrithiamine were also found to have mutations in the genetic region coding for the TPP riboswitch, and X-ray diffraction crystal structures of the TPP riboswitch in complex with PTPP revealed a near identical binding mode to TPP.[107,108] The TPP riboswitch is also a promising target due to the fact that several fragment hits have been verified, which is a feature commonly associated with druggable binding sites.[109,110]

Another example of a riboswitch with a library of known binders is the guanine riboswitch. Hypoxanthine, a metabolite and a close analogue of guanine was found to selectively bind to the guanine riboswitch over the adenine riboswitch, and a crystal structure of the complex with 1.95 Å has been reported.[111] Several modified guanine structures with various modifications were developed and found to be active in an in-line probing assay, with K_D values even exceeding that of the natural ligand (0.5 nM *vs.* 5 nM for guanine).[112] Other analogs of guanine were found to modulate the guanine riboswitch and successfully display bactericidal activity at relatively low concentrations.[113]

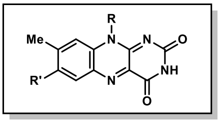
L-Aminoethylcysteine and DL-4-oxalysine were originally reported in the 1950s and 60s as analogs of lysine which were effective in inhibiting the growth of specific Gram-positive bacteria.[114,115] More recent research has shown that these compounds are able to bind to the lysine riboswitch, but also that bacteria with resistance towards these compounds have mutations in the genome that encodes for this specific riboswitch.[116,117]

1.8 The FMN riboswitch

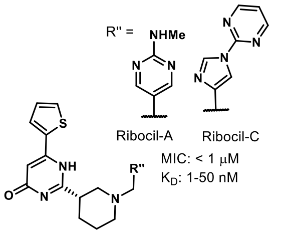
One of the most studies riboswitches is the FMN riboswitch, which is responsible for the expression of genes involved in biosynthesis and transport of riboflavin.[118–123] Riboflavin plays a number of roles, such as in oxidative phosphorylation and energy transfer in the citric acid cycle. While essential in both mammalian and bacterial biology, an important difference lies in the acquisition of this molecule: mammalians harvest the riboflavin vitamin through diet, while all bacterial cells are self-sufficient, and synthesize this molecule.[124] With this adaptation, the need to self-regulate the biosynthesis and maintain homeostasis of riboflavin arises, which has been addressed through the FMN riboswitch. The natural ligand to this riboswitch is the phosphorylated version of this biomolecule, flavin mononucleotide (FMN, **Table 1**). A strong case for exploring the FMN riboswitch as a potential new antibiotic target, is its widespread occurrence and conservation in bacteria, appearing in 41 human pathogenic species, where 7 are on the list of the WHO's priority pathogen list.[102,103] Seemingly all Gram-negative bacterial species rely on a single riboswitch to regulate *de novo* synthesis, while some Gram-positive bacteria, such as *Enterococcus faecalis* and *Streptococcus pyogenes*, which are riboflavin auxotrophs, still rely on a FMN riboswitch to transport riboflavin from its host environment.[125]

The FMN riboswitch has a plethora of known ligands with high affinity other than FMN, some of which are shown in **Table 1.** As with the TPP riboswitch, the FMN riboswitch discriminates strongly between the phosphorylated and the unphosphorylated variants of its natural ligand.[126] Roseoflavin (**RoF**), a naturally occurring antimicrobial agent isolated from *Streptomyces davawensis*, is a structural analog to FMN, with the main difference in the exchange of a methyl group with a dimethylamine group and the lack of a phosphate group.

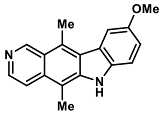
Table 1: Some ligands for the FMN riboswitch and their affinities



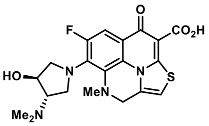
Riboflavin scaffold



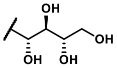
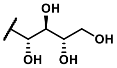
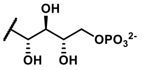
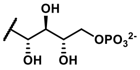
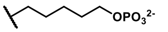
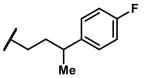
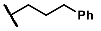
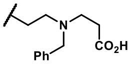
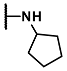
MIC: < 1 μM
K_D: 1-50 nM



WG-1
MIC: 3.1 μM
K_D: ND



WG-3
MIC: <0.1 μM
K_D: ~130 nM

Compound	R =	R' =	K_D	Compound	R =	R' =	K_D
Riboflavin		Me	~39 μM	Roseoflavin (RoF)		NMe_2	~0.2-1 μM
FMN		Me	~1-37 nM	RoFMN		NMe_2	~3 nM
BRX830		NH_2	~2.6 nM	5FDQD		NMe_2	~7.5 nM
BRX1555		Me	~39 nM	BRX1354			~16 nM

The phosphorylated variant (RoFMN) has been shown to inhibit the growth of bacteria and RoF-resistant strains of *B. subtilis* and *Lactococcus lactis* have been identified with mutations in the FMN riboswitches which are associated with *ribD* (an enzyme in the riboflavin biosynthesis pathway), strongly indicating binding in the same binding pocket as FMN.[127,128] It was later verified through X-ray diffraction that this compound indeed binds to the FMN riboswitch.[99,122] Exploring further variations on the riboflavin structure, 5FDQD (**Table 1**) was identified as a synthetic flavin analog which also displayed high potency towards the FMN riboswitch and an ability to cure mice with laboratory-induced *C. difficile* infections.[129] Lead optimization led to further analogs, such as BRX830, BRX1151 and BRX1354 (**Table 1**), which have various substitutions and allowed to establish structure-activity research (SAR)-principles for the FMN riboswitch.[130] Ribocil A and ribocil C (**Table 1)** were serendipitously discovered by Merck in a phenotypic screening campaign for compounds active in disrupting the riboflavin biosynthetic pathway in *E. coli*.[131,132] Mapping the genome of the ribocil-resistant bacterial strains revealed mutations in the region corresponding to the FMN riboswitch. The ribocils were found to not only bind to the riboswitch with high specificity, but also displayed higher antimicrobial potency than RoF (<1 *vs.* >128 µg mL^{-1}). X-ray structures of ribocil A and an analogue in complex with the target have been published.[131,132] The ribocils stands as a turning point in the development of ligands for the FMN riboswitch, as its scaffold departs from the riboflavin-like scaffold found in the previously identified binders. In addition to the ribocils, WG-1 and WG-3 were discovered through a mass spectrometry-based affinity screen using an Automated Ligand Detection System (**ALIS**), based on the same compound library used in Merck's previous screening campaign.

1.9 Binding mode of FMN and other ligands

The binding mode of FMN in the FMN riboswitch is shown in **Figure 10**. The flavin ring is sandwiched between the nucleobases of A85 and A48, which provides significant hydrophobic interactions through π-π stacking interactions. Hydrogen bonds are formed between one of the carbonyl moieties of FMN and two nearby residues: the sugar backbone of A48 and the exocyclic nitrogen of A99. An endocyclic nitrogen atom of FMN also forms another hydrogen bond to the 1N atom of A99.

Roseoflavin, 5FDQD and the BRX-series (**Table 1**) of derivatives all share the flavin scaffold, and thus adopt an identical binding mode with respect to this moiety. Several polar contacts are formed between FMN's phosphate group and the binding site (not shown). Although the latter contacts appear to contribute significantly to the binding affinities of FMN and riboflavin, the potencies of 5FDQD and BRX1555 clearly show that also compounds that bear a neutral group in this position can bind potently. As previously discussed, the ribocils depart from the flavin scaffold while still showing high activity. In the crystal structure obtained of the riboswitch/ribocil A complex (**Figure 10C**), the same π-π sandwiching between A48 and A85 is observed, with two hydrogen bonds between the pyrimidinone carbonyl of ribocil and the A48 backbone and A99 also conserved. Ribocil A also makes two new contacts with the binding site which notably differs from FMN: a π-π stacking interaction between the aminopyrimidine of ribocil A and residue G62 and an edge-face π -interaction between the thiophene moiety of ribocil A and A49.

.

Previous work in the Brenk group has involved virtual screenings of commercially available compounds to find novel ligands for the FMN riboswitch. One of the lead compounds, **RSL-0035** (**Figure 11C**) showed promising activity in an isothermal calorimetric (ITC) displacement assay, where it was shown that the presence of the compound successfully weakens the affinity of FMN to the FMN riboswitch, and it was estimated to have a K_D around 240 µM.[133] Binding was also detected in a surface

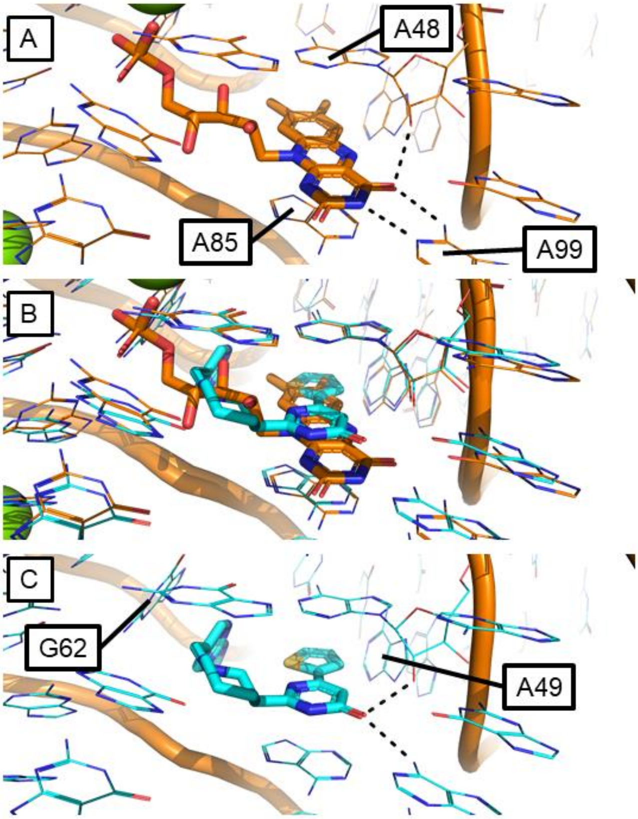


Figure 10: X-ray diffraction crystal structure showing the binding modes of the ligands (thick sticks) FMN (orange) and ribocil (cyan) in the FMN riboswitch (thin sticks: residues, tubes: backbone. Hydrogen bonds shown as black dashes **A)** The FMN riboswitch bound to FMN. PDB ID: 2yie **B)** FMN and ribocil superimposed **C)** The FMN riboswitch bound to ribocil, PDB ID: 5c45

plasmon resonance (SPR) assay at a single concentration, but it was not possible to determine a K_D.

RSL-0035 shares many similarities with FMN: Both compounds are composed of three fused aromatic rings (A, B and C, **Figure 11**) where ring A resembles uracil. A notable difference is the kink caused by the five-membered B ring in **RSL-0035**. The predicted binding mode of **RSL-0035** in the FMN riboswitch binding site after molecular minimization (**Figure 11**) suggests that the three rings of **RSL-0035** are sandwiched between A48 and A85, and rhe A ring forms base pairs with A99 and additional hydrogen bonding with the backbone hydroxyl group of A48. The modelling did not show any significant interactions from the phenyl group in the 9-position in ring C, nor from the methyl group in position 7. This compound stands as a promising and novel ligand for the FMN riboswitch with potential for optimization and has formed the basis for the ligand design and synthesis work of this thesis.

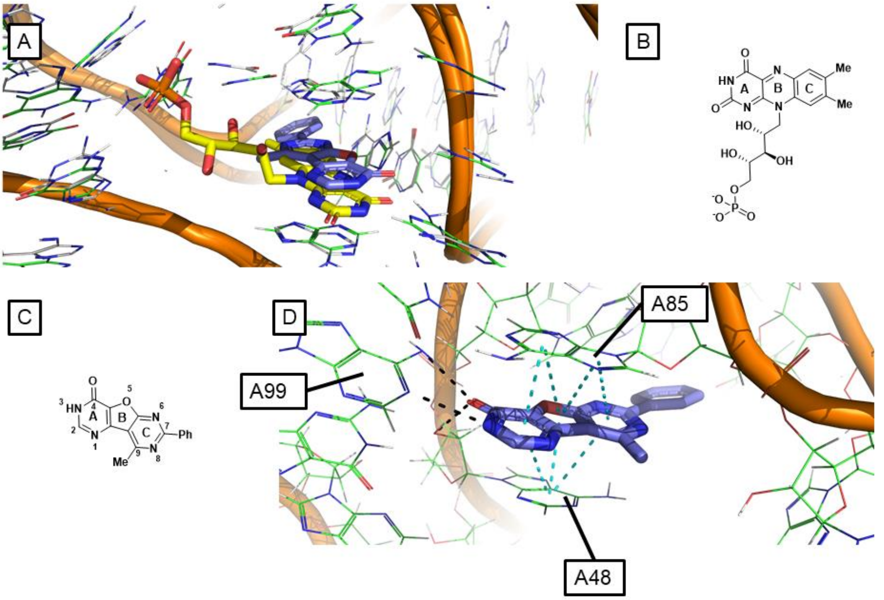


Figure 11: A) X-ray diffraction crystal structure (PDB ID: 2yie) of the FMN riboswitch (grey thin sticks) complexed with FMN (thick yellow sticks) superimposed with the minimized structure of **RSL-0035** (blue thick sticks) and the FMN riboswitch (thin green sticks). Backbone shown as orange tubes **B)** Structure of FMN **C)** Structure and labeling of **RSL-0035** atoms **D)** Rotated view of the minimized structure of the FMN riboswitch complexed with **RSL-0035**, highlighting the molecular interactions with the binding site. Hydrogen bonds shown as black dashes, π-π interactions shown as cyan dashes.

2. Aims

The overall aim of this thesis has been to explore the ligand space of riboswitches to find potential lead compounds for future antibiotics. In this work we have put special emphasis on the FMN riboswitch as a drug target but have also characterized other RNA binding sites with respect to druggability. The main goal has been addressed through the following sub-goals:

- Develop an open access druggability predictor (based on the framework of DrugPred 2.0) which is compatible with RNA crystal structures, evaluate and validate its performance on both protein and RNA binding sites, and discuss its performance on the available RNA crystal structures in the Protein Data Bank.
- Explore the ligand space of the FMN riboswitch and propose ligands which can act as novel binders to this target.
- Resynthesize the virtual screening hit **RSL-0035** to validate its structure and affinity towards the FMN riboswitch in an activity assay.
- Synthesize two series of proposed FMN ligands and suggest new ligands based on obtained structure-activity relationship (SAR) data.

3. Results and discussion

3.1 DrugPred_RNA

The results presented in this thesis related to DrugPred_RNA formed the basis of a publication in the *Journal of Chemical Information and Modeling* in 2021. As such, the figures and tables are taken from this publication and the text has been modified to fit within the thesis.[71]

3.1.1 Construction of DrugPred_RNA

As has been discussed in the introduction, RNA is a promising target. However, we lack a tool that allows identification of druggable pockets for prioritization. Therefore, the goal of this part of the thesis was to adopt DrugPred to RNA pockets. As mentioned, a druggability analysis of RNA crystal structures has been performed using DLID.[87] This study concluded that several RNA binding sites are suitable for drug development, but the method does not clearly discriminate between ligandable sites (in a broader sense). DrugPred, a previously developed druggability predictor for protein binding sites in our group, should easily be converted to be applicable for RNA crystal structures. The underlying principles of what defines druggable from less druggable binding sites should hold true for both protein and RNA binding sites, and as there is a dearth of information about either druggable or less druggable RNA binding sites, a prediction software compatible with RNA should be rather trained on protein binding sites.

The following minor adjustments were made in construction of the new version of DrugPred, from here on called DrugPred_RNA: 1) Descriptor calculations were made using RDKit, an open-source cheminformatics package, opposed to the licensed OpenEye software used in DrugPred 2.0, the last iteration of the software 2) Some descriptors used in DrugPred 2.0 were specific for amino acids (such as hydrophobicity indices), and incompatible with RNA binding sites. Therefore, some descriptors were removed, and some new descriptors were calculated for DrugPred_RNA. 3) Initial attempts at using PLS-DA, the machine learning (ML) method used in DrugPred 2.0, failed to give satisfactory predictive power. A different ML method, XGBoost (eXtreme Gradient Boosting), was used in DrugPred_RNA instead.[134]

To define the boundaries of the binding sites, a library of small molecules was docked in the binding sites after the cognate ligands were removed. The atoms of molecules which scored above a certain energy threshold were merged into one common structure to act as a negative mold of the binding site, hereafter referred to as a superligand. In total, 23 descriptors were calculated for the construction of DrugPred_RNA (**Table 2**), which overall describe geometrical and chemical properties of each ligand binding site. Details and specifics of how these descriptors were obtained are described in more detail within the methods part of this thesis.

With the descriptors calculated for all the protein bindings sites found in the NRDLD, a new predictor was trained using the same training/testing split used in the construction of the last iteration of DrugPred. Overfitting was avoided by fine tuning of XGBoost settings, limiting the maximum depth of trees and using an early stopping round option. Analysis using Shapley addictive explenations (SHAP) values of the model's performance proved to be a valuable tool to eliminate a "black box" situation and assure that the descriptors with highest impact on the model's output were consistent with theory by ranking each descriptor's predictive impact.[135] Removing the descriptors with the lowest impact on the model's output in a step-by-step method until the model's output was negatively affected, a predictive model was created which used 12 of the 23 originally calculated descriptors (**Figure 12A**).

Table 2: List of descriptors describing size, polarity and shape of ligand binding sites used in the construction of DrugPred_RNA. Descriptors which were used in the final version of DrugPred_RNA, are denoted with a mean absolute SHAP value.

Descriptor name	Description	Descriptor type	Mean absolute SHAP value
csa	Sum of SASA of binding site atoms	size	NA
psa	Sum of SASA of polar binding site atoms	polarity	NA
psa_r	psa / csa	polarity	1.46
hsa	Sum of SASA of hydrophobic binding site atoms	polarity/size	0.30
ali_sa_r	Sum of SASA of aliphatic binding site atoms / *csa*	polarity	NA
exp_sl_sa	Sum of SASA of superligand atoms that are solvent exposed in the superligand-receptor complex	shape	0.224
no_sl_atoms	Number of superligand atoms	size	0.202
no_bs_atoms	Number of binding site atoms	size	0.167
fr_buried_sl_atoms	Number of atoms buried inside the superligand / number of superligand surface atoms	shape	0.345
fr_hpb_atoms	Number of hydrophobic binding site atoms / no_bs_atoms	polarity	0.629
sl_bs_r	no_sl_atoms / no_bs_atoms	shape	0.193
vol	Volume of superligand	size	NA
sa_vol_r	Surface area of superligand / *vol*	shape	0.0907

PMI1	First principal moment of inertia	shape/size	NA
PMI2	Second principal moment of inertia	shape/size	NA
PMI3	Third principal moment of inertia	shape/size	0.277
NPR1	PM1 /PM3	shape	NA
NPR2	PM2 / PM3	shape	NA
Asphericity	$0.5\frac{(PM3-PM1)^2+(PM3-PM1)^2+(PM2-PM1)^2}{PM1^2+PM2^2+PM3^2}$	shape	NA
Eccentricity	$\frac{\sqrt{PM3^2-PM1^2}}{PM3^2}$	shape	NA
SpherocityIndex	$\frac{3\times PM1}{(PM1+PM2+PM3)}$	shape	0.0825
RadiusOfGyration	$\sqrt{\frac{2\pi^{\frac{PM3\times PM2\times PM1}{3}}}{MW}}$	shape	NA
InertialShapeFactor	$\frac{PM2}{PM1\times PM3}$	shape	0.0849

3.1.2 Initial evaluation of DrugPred_RNA

The descriptor with the highest impact was found to be *psa_r*, which together with *fr_hpb_atoms* (ranked second) and *hsa* (ranked fourth) describe the polarity/hydrophobicity of the binding sites. As expected, binding sites which were more polar were more likely to be less druggable and *vice versa* (**Figure 12B**). The remaining 9 descriptors describe geometrical properties of the binding site, with *fr_buried_sl_atoms,* describing the compactness of the binding site, having the highest impact (ranked third). It is not surprising to see that polarity plays overall a more

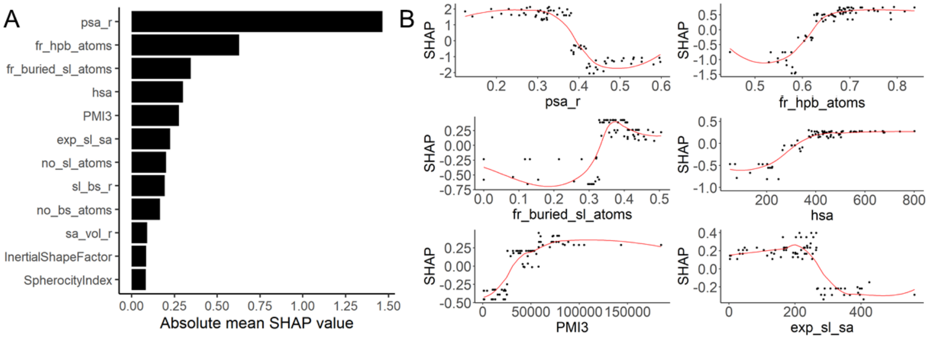


Figure 12: SHAP values for the DrugPred_RNA model. A) Absolute mean SHAP values for each descriptor ranked from the highest to lowest impact on the model output. B) Individual SHAP values for each pocket in the training set for the top six descriptors in the model plotted against the descriptor values. Locally estimated scatterplot smoothing (LOESS) curves are overlaid on the descriptor observations (black dots). The midpoint in each curve indicates the cutoff value from where the prediction changes the direction. Positive SHAP values are associated with druggable and negative SHAP values with less druggable binding sites.

important role than shape, but what was found surprising is that none of the descriptors which were designed with size in mind (*csa, vol, no_sl_atoms, no_bs_atoms*) were found to be of significant value or to have any predictive importance, with *no_sl_atoms* highest at the seventh rank. It is likely the hydrophobic surface area (*hsa*) plays a dual role in binding site description; larger binding sites are also expected to have larger hydrophobic surface areas.

The predictor had been trained on the same training/testing-set with the NRDLD which the previous iteration of DrugPred was based on, and its performance is shown in . The performance of DrugPred_RNA is comparable or better to DrugPred 2.0. In the final model, 1 of the 75 binding sites in the training set was misclassified as less druggable (**Figure 13**), and within the testing set, 4 of the 35 binding sites were misclassified (2 false positives and 2 false negatives).

Table 3: Performance of DrugPred_RNA and DrugPred 2.0 on the training and testing set of the NRDLD

	Training set [druggable / less druggable]		Test set [druggable / less druggable]	
	DrugPred_RNA	DrugPred 2.0	DrugPred_RNA	DrugPred 2.0
Accuracy	0.99	0.91	0.91	0.94
Precision	1.00 / 0.97	0.92 / 0.89	0.95 / 0.86	0.95 / 0.93
Recall	0.98 / 1.00	0.94 / 0.86	0.91 / 0.92	0.95 / 0.93

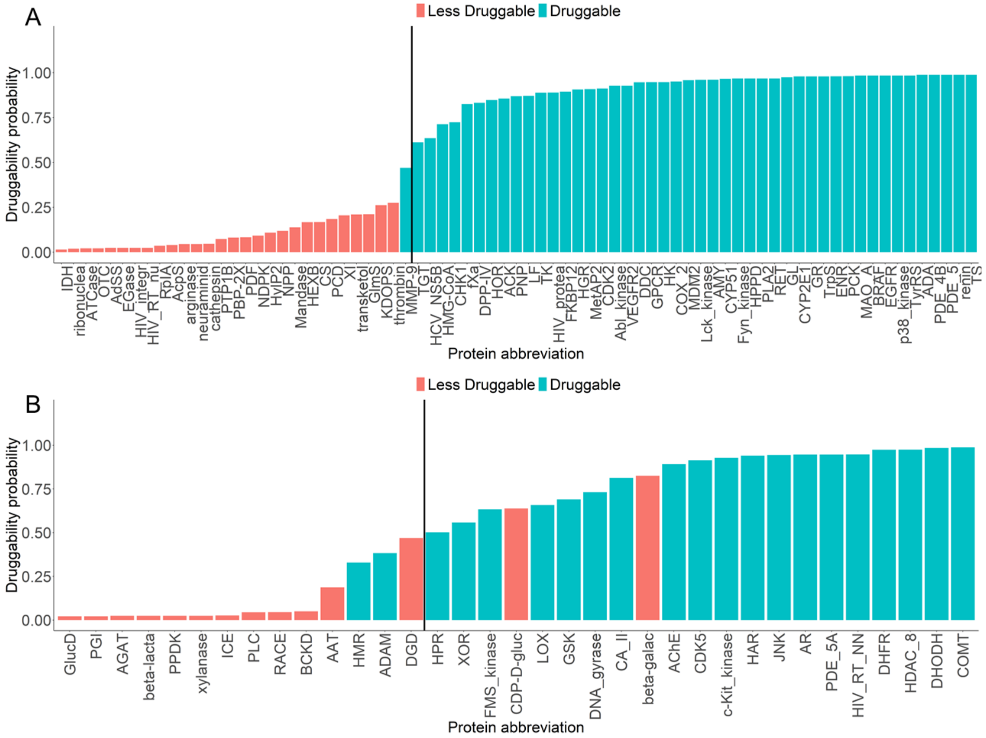


Figure 13: Druggability predictions with DrugPred_RNA for the NRDLD training **(A)** and test set **(B)**. Cyan bars represent druggable and red bars less druggable binding sites. All pockets at the right side of the black line are classified as druggable while the pockets on the left side of the line are classified as less druggable. The full names of the proteins are listed in the publication related to DrugPred 2.0.[3]

3.1.3 Prediction on RNA binding sites

After the model had been created, it was ready to be applied on RNA-containing binding sites. Two data sets of crystal structures were compiled from the Protein Data Bank (PDB), one with RNA-exclusive binding sites, and one composed of ribosomal binding sites, which could contain a mixture of ribosomal RNA and protein macromolecules (**Table 4**). To investigate the effect of metals on the binding sites druggabilities, copies were made of binding sites with metal ions within 5Å of the ligand with the metal removed. The RNA dataset contained 427 unique PDB entries spanning 465 metal-free binding sites and 343 metal-containing binding sites, with 217 distinctive ligands. The ribosomal set spanned 497 PDB entries with 613 metal-free pocket and 546 metal-containing pockets, with a total of 217 unique ligands. Both datasets had comparable druggability instances, with 36% and 31% of the RNA and the ribosomal set containing druggable binding sites, respectively. To assess the robustness of the model, prediction scores of similar binding sites were compared by grouping them together through the alignment of binding site similarities. Each family consisted of binding site sequences with >85% similarity, and each family was given a consensus score using the following equation:

$$C = \frac{|n_d - n_{ld}|}{N} \times 100\%$$

Where C is the consensus score, n_d is the number of druggable binding sites within a family, n_{ld} is the number of less druggable binding sites, N is the number of binding sites within the same family. For example, if all members within a family scored druggable/less druggable, a consensus score of 100% was obtained. If half scored druggable and the other half less druggable, a consensus score of 0% was obtained.

Table 4: Data Sets of RNA and Ribosomal Binding Sites for Assessing DrugPred_RNA

	RNA-only set (metal free/metal-containing set)	ribosome set (metal free/metal-contining set)
Unique PDB IDs	427	497
Binding sites containing small molecule ligands	465/343	613/546
Unique ligands	224	217
Druggable entries	172/126	224/141

A histogram of consensus scores for both datasets are shown in **Figure 14**. As can be seen, consensus scores between 90 and 100% is the most prevalent of the families in both datasets, showing that DrugPred_RNA provides reliable predictions on similar binding sites. The presence of a metal in a binding site was also evaluated, and as can be seen, the consensus scores do not seem to be affected in a significant way. For 90% of all binding sites in the RNA dataset and 83% in the ribosome set, the druggability predictions were unaffected by the presence of metal ions. Thus, only metal free binding sites are discussed hereafter. The result on robustness is discussed below.

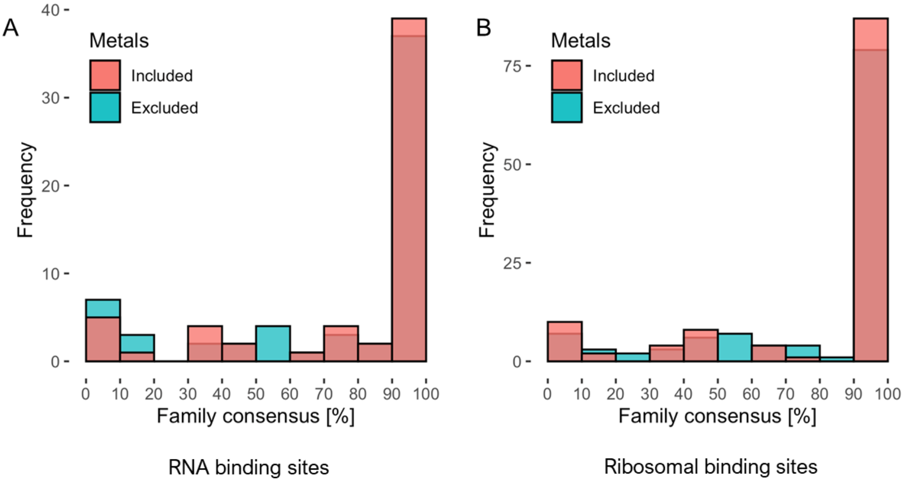


Figure 14: Consensus scores obtained for the two datasets assessed by DrugPred_RNA.

The distribution of the descriptors calculated for the binding sites in both the NRDLD and in the RNA and ribosomal dataset are shown in **Figure 15**. In general, the descriptors for the druggable protein binding sites were more narrowly distributed than those for less druggable protein binding sites or the RNA pockets. Both RNA sets (ribosomal and RNA-only) had binding sites for which the descriptor values were in the same range as those found for druggable protein pockets.

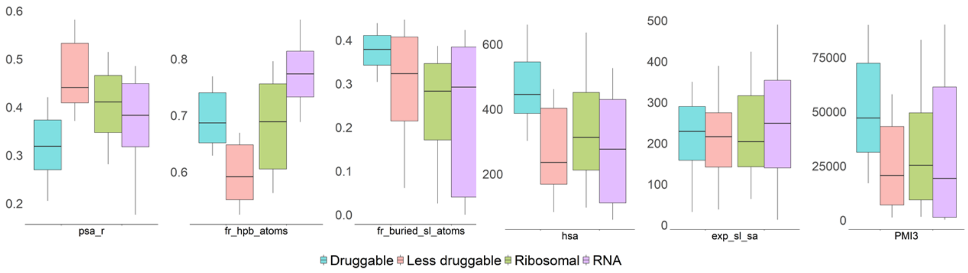


Figure 15: Boxplots showing the distribution of the six highest descriptors in the DrugPred_RNA model. Druggable and less druggable refer to the protein binding sites in the training and testing set used in the construction of DrugPred_RNA, ribosomal and RNA refers to the two datasets the finished model was applied on**.**

3.1.4 Evaluation of predictions of DrugPred_RNA on RNA binding sites

The performance of DrugPred_RNA was then further evaluated on the basis of following criteria:

1. The agreement of druggability predictions and visual inspections of binding sites
2. The extent to which binding sites that efficiently bind drug-like ligands were predicted to be druggable
3. The extent to which drug-like ligands efficiently bind to binding sites predicted to be druggable, and the robustness of the predictions with respect to substitutions and conformational changes in the binding sites

3.1.5 Visual inspection and manual assignment of binding sites.

As the realm of RNA drug targets is underexplored, a few select examples from the RNA set are chosen to highlight the binding sites with known ligands as a first validation of the predictions. The examples were selected to have published affinity data for at least the cocrystallized ligand, cover different RNA classes, and have different prediction outcomes. Two ribosomal pockets, three riboswitch pockets, a trans-activation response (TAR) element RNA and a splicing site were included. The

SHAP values show that pockets which were large and hydrophobic and able to properly envelope the ligand were deemed druggable by DrugPred_RNA, which is confirmed after visual analysis of the binding sites. These include the binding site of linezolid (**Figure 16A**), the FMN riboswitch **(Figure 16B**) and a TAR RNA binding site (**Figure 16C**). Bindings sites which are small and/or unable to provide a significant hydrophobic contact surface would typically be deemed less druggable. DrugPred_RNA agreed with this, and as such finds the guanine (**Figure 16D**) and the lysine riboswitch (**Figure 16E**) binding pockets to be less druggable., along with that of a binding site of a splicing site modifier (**Figure 16F**) and the binding site of paromomycin.

3.1.6 Relation between drug-like ligands and corresponding binding site prediction

In the next evaluation step, binding sites which contained drug-like ligands were examined, with the expectation that the binding sites of highly drug-like molecules are predicted to be druggable.[52,53] In the two data sets, 18 of 331 ligand had a QED score ≥ 0.67, which is the benchmark score for highly drug-like molecules. 12 of these were found in druggable pockets (67%), while 6 were found in binding sites assessed to be less druggable (23%). which when compared to the druggability predictions for all metal-free binding sites in the two combined datasets is a considerable enrichment of the pockets predicted to be druggable (37%). As only 37% of all metal-free binding sites were predicted to be druggable, the drug-like ligands were clearly enriched in druggable binding sites.

For the next evaluation we looked at the relationship between ligands which bind tightly and predictions of pockets. In this context, we considered a ligand to bind tightly to a binding site if it had a ligand efficiency (the binding energy normalized by the number of heavy atoms, LE) at least close to 0.30 kcal· mol^{-1} ·heavy atom^{-1}, which translates to low nanomolar binding affinities of compounds with a molecular weight of maximum 500 Da under the assumption that the ligand efficiency stays at its best constant during optimization.[136] For 10 out of the 12 drug-like ligands binding to pockets predicted to be druggable, we could find binding data in the literature (**Table 5**). Based on these data, eight ligands bind efficiently to their target with ligand efficiencies (LEs) > 0.30 kcal·mol^{-1}·heavy atom^{-1}, hinting that these pockets are indeed druggable. The two remaining ligands were linezolid with the 50S ribosomal subunit as target and acetylpromazine binding to HIV-1 TAR RNA (**Figure 16A, C**). Based on manual assignment (see the supplementary material in the related publication), these pockets also appear to be druggable. Thus, all predictions for the pockets binding to the 10 drug-like ligands with accessible binding data appear to be valid.

Table 5: Drug-like ligands (QED ≥ 0.67) found in RNA binding sites predicted to be druggable.

Ligand ID	**PDB ID**	**Receptor name**	**QED score**	**K_D [nM]**	**LE [kcal·mol^{-1}·heavy atom^{-1}]**
RNA data set					
MGR	1q8n	Malachite green aptamer	0.76	800[137]	0.34
6YG	5kx9	FMN riboswitch	0.69	13.4[132]	0.41
L8H	2l8h	HIV-1 TAR RNA	0.67	NA#[138]	-
PMZ	1lvj	HIV-1 TAR RNA	0.85	27,000[139]	0.22
Ribosomal data set					
917	5v7q	50S ribosomal subunit	0.94	700[140]	0.39
ZLD	3cpw	50S ribosomal subunit	0.89	20,000[141]	0.27
G6M	6ddg	50S ribosomal subunit	0.79	2,600[142,143]	0.31
3HE	4u3u	80S ribosome	0.76	140[144]	0.48
G6V	6ddd	50S ribosomal subunit	0.76	2,600[142]	0.30
ANM	3cc4	50s ribosomal subunit	0.78	20,000[145]	0.34
HN8	5on6	80S ribosome	0.71	NA#	-
3K8	4u55	80S ribosome	0.71	39	0.32

# binding affinity unknown

On the other hand, six drug-like ligands were found in pockets predicted to be less druggable (**Table 6**). For five of them, we could retrieve affinity data in the literature, and all of these bind rather efficiently to their targets (LE $\geq$ 0.29 kcal$\cdot$ mol^{-1} $\cdot$heavy atom^{-1}). Three of these ligands are fragments binding to the TPP riboswitch, one is a ligand binding the influenza A virus promoter region, and one a ligand of the Spinach aptamer. Several examples of the TPP riboswitch binding site were contained in the RNA-only set (**Figure 18**). The pockets differ mainly in the conformation of G72 (**Figure 18E),** but in all cases, the pocket is rather large and partially buried (**Figure 18A-D).** The pockets with G72 in one of the conformations were predicted to be druggable (**Figure 18A, B**), while pockets with G72 in the alternative conformation (**Figure 18C, D**), including the ones binding the drug-like fragments, were predicted to be less druggable. Based on the structures, discussed in more detail below, it is not obvious why the latter TPP riboswitch binding sites should be less druggable

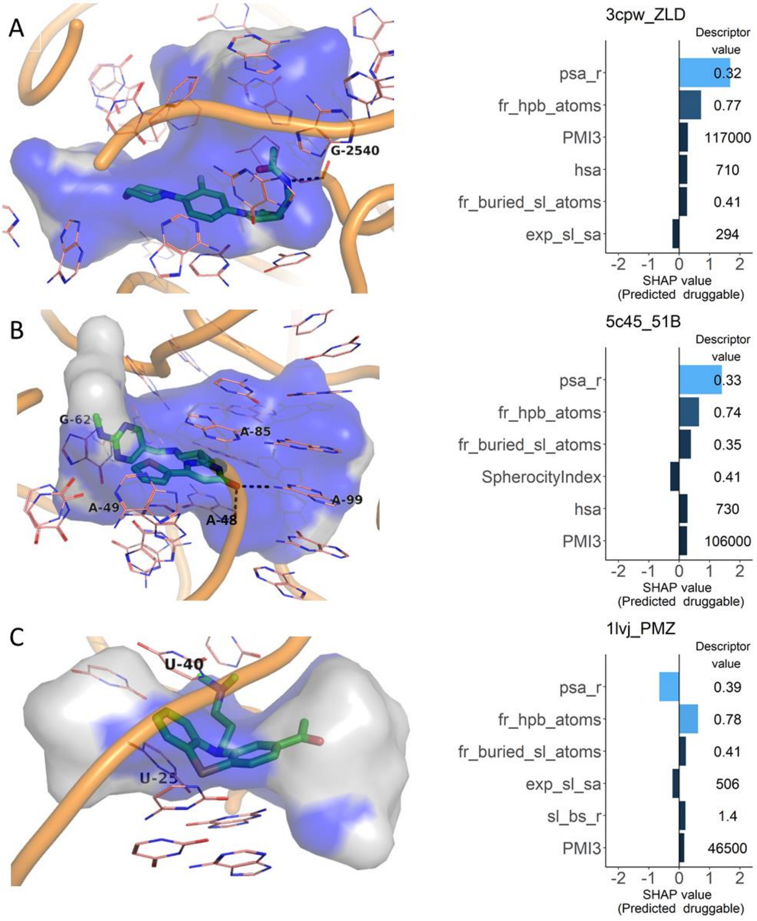
Table 6. Drug-like ligands (QED ≥ 0.67) found in RNA binding sites predicted to be less druggable.

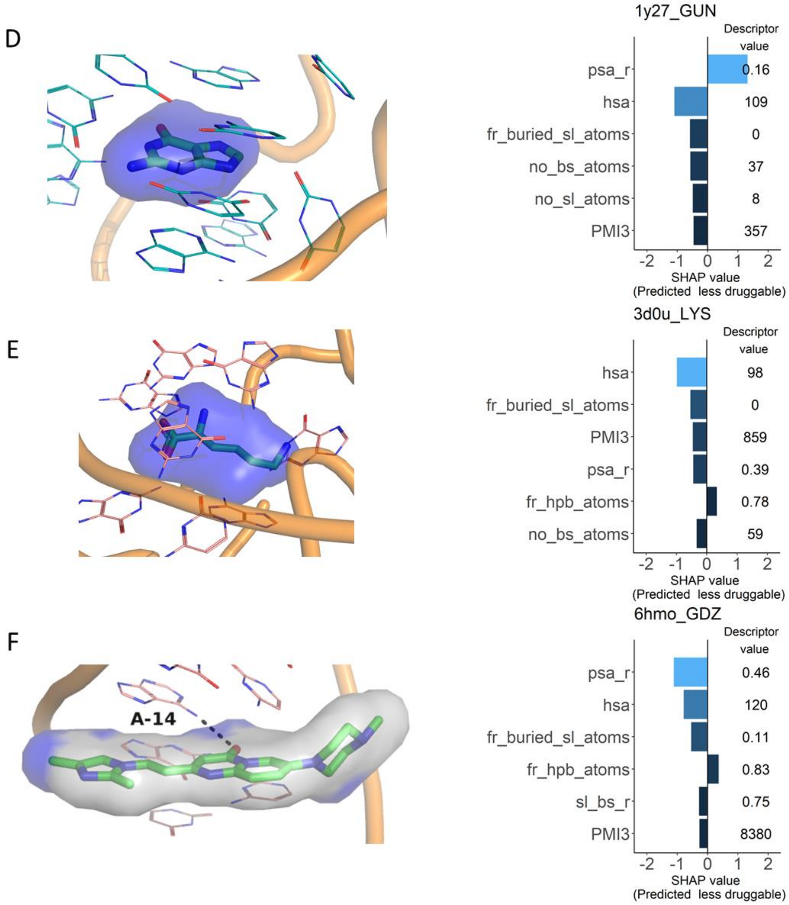
Ligand ID	**PDB ID**	**Receptor name**	**QED**	**K_D [nM]**	**LE [kcal·mol^{-1}·heavy atom^{-1}]**
RNA data set					
VIB	4nyg	TPP riboswitch	0.79	1,500[110]	0.45
2QC	4nyb	TPP riboswitch	0.77	103,000[110]	0.43
0EC	2lwk	Influenza A	0.86	50,000[146]	0.29
1TU	5ob3	Spinach aptamer	0.85	530[147]	0.49
218	2hop	TPP riboswitch	0.77	6,000[148]	0.38
Ribosomal data set					
TRP	4v6o	Tryptophan-sensing ribosomal site	0.67	NA#	-

# binding affinity unknown

These predictions can therefore be considered false negative. The drug-like ligand of the influenza A promoter region sits on the surface of the RNA molecule and is almost entirely solvent exposed (**Figure 17A**). It is highly unusual that a ligand with such a binding mode binds that efficiently (LE = 0.29 kcal·mol^{-1} ·heavy atom^{-1}). However, the structure of the complex has been determined by NMR, and it is possible that the resolution of the structure is not accurate enough to reveal the details of the binding mode.[146] The small molecule dye, DFHBI, is bound deep into the solvent-excluded part of the pocket in the Spinach aptamer, forming pi-stacking interactions and hydrogen bonds with the surrounding residues (**Figure 17B**). Considering the drug-likeness of the ligand together with its efficient binding and its binding mode, the prediction for this pocket by DrugPred_RNA is likely wrong.

The druggability predictions for the pockets predicted to be druggable and binding to drug-like ligands appeared to be correct, while the highly drug-like ligands which were found in less druggable pockets were found to be either false negative due to edge cases of conformational variety, likely incorrect structures and only one case of a clear false assignment from method. These results suggest that DrugPred_RNA is more likely to assign druggable sites as less druggable than *vice versa*, which is also observed in the results of the NRDLD test, but the data points discussed in this paragraph are too scarce to make a definitive conclusion.




D
1y27_GUN
Descriptor value
psa_r
hsa
fr_buried_sl_atoms
no_bs_atoms
no_sl_atoms
PMI3
0.16
109
0
37
8
357
-2
-1
0
1
2
SHAP value
(Predicted less druggable)
E
3d0u_LYS
Descriptor value
hsa
fr_buried_sl_atoms
PMI3
psa_r
fr_hpb_atoms
no_bs_atoms
98
0
859
0.39
0.78
59
-2
-1
0
1
2
SHAP value
(Predicted less druggable)
F
A-14
6hmo_GDZ
Descriptor value
psa_r
hsa
fr_buried_sl_atoms
fr_hpb_atoms
sl_bs_r
PMI3
0.46
120
0.11
0.83
0.75
8380
-2
-1
0
1
2
SHAP value
(Predicted less druggable)

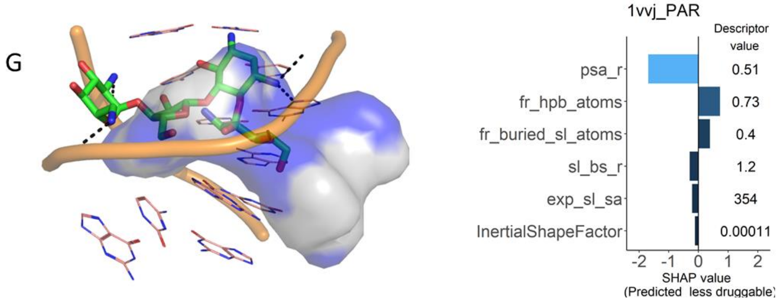


Figure 16: Evaluation of the performance of DrugPred_RNA based on selected examples. The RNA backbones are shown as orange tubes, nucleobases as thin sticks with carbon atoms colored pink, and ligands as thick sticks with carbon atoms in green. The surface of the superligand created by DrugPred_RNA as a negative print of the pocket is shown as blobs with the solvent exposed surface area colored gray and the remaining surface area colored blue. Hydrogen bonds are indicated as dotted black lines. For each pocket, the individual SHAP values for the six most important descriptors together with the descriptor values are also displayed. The SHAP value plots are labeled with the PDB IDs of the receptors and the three-letter codes of the ligands found in each pocket. **(A)** The binding site of linezolid in the 50S ribosomal subunit. **(B)** Ribocil A bound to the FMN riboswitch. **(C)** TAR RNA complexed with acetylpromazine. **(D)** Guanine bound to the guanine riboswitch. **(E)** Lysine in the binding site of the lysine riboswitch. (F) Splicing site complexed with a splicing site modifier. **(G)** Paromomycin bound to a bacterial ribosome site.

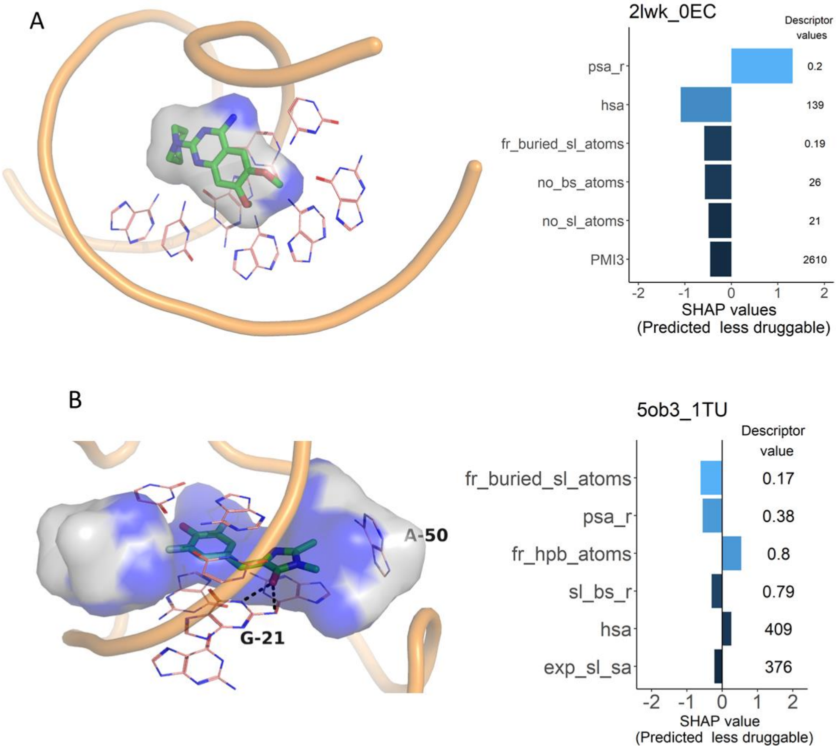


Figure 17. RNA binding sites predicted to be less druggable, but binding drug-like ligands. The surface of the superligand created by DrugPred_RNA as a negative print of the pocket is shown as a blob with the solvent exposed surface area coloured grey and the remaining surface area coloured blue. For each pocket, the individual SHAP values for the six most important descriptors together with the descriptor values are also displayed. The SHAP value plots are labelled with the PDB IDs of the receptors and the three letter codes of the ligands found in each pocket. A) Binder (green) of influenza A promoter region (PDB ID 2lwk). B) The Spinach aptamer (PDB ID 5ob3) bound to the dye DFHBI (green).

3.1.7 Robustness of DrugPred_RNA

As mentioned above, binding sites were grouped in families based on binding site sequence similarity. As the binding site sequences are rather short (on average, about 15 residues for the RNA data set and 47 for ribosomal data set), a low cutoff score (>85%) was set to allow some variation within the binding site sequence. Similar sequences were then grouped together, and each group was given a unique family ID. Robustness was also assessed on families which were grouped together based on their global sequence alignment, where a higher cutoff value was used (>98%). This grouping was only done for the RNA data sets, as ribosomes often contain multiple binding sites, which would have resulted in different pockets within the same structure assigned to the same family. In the RNA set, 57 families were assigned on global sequence similarity, and 46 based on binding site similarity (Table S 1 in the Appendix of this thesis). In the ribosomal set, 52 families were assigned from binding site sequence similarity (Table S 2). Most of the families in the RNA dataset obtained a 100% consensus score (79% for global sequence similarity, 74% for binding site similarity). In the ribosomal dataset 75% of the binding sites obtained the same prediction score. As such, in most cases, different crystal structures of the same binding site were predicted with the same outcome.

Some examples of low consensus-scoring families were explored to elucidate what could cause such inconsistencies. The first example is the TPP riboswitch family, which had binding sites from 16 distinct PDB entries. The consensus score for the binding site sequenced family here was 13%, with a majority predicted to be less druggable. This was surprising, as the TPP riboswitch has been earlier expected to be a good drug target candidate. Superimposing all the binding sites revealed that the binding was subject to some plasticity, specifically from a particular guanine residue (G72 in the *E. coli* riboswitch), which adopts different conformations from crystal structure to crystal structure (**Figure 18A, C, E**). The resulting superligands are rather uniform in polarity, but vastly different in shape and size, leading to disparaging druggability predictions (**Figure 18B, D**). When bound to TPP, this residue stays more consistently in one conformation. From inspecting the structures which have been predicted to be less druggable, there is no clear reason why they should have been predicted to be less druggable. Considering this and the affinities of the ligands, it is likely DrugPred_RNA has provided false negatives in these examples.

Another example of a family with a low consensus score is the ZNP riboswitch, which was composed of three entries in total, all bound to the same ligand, ZMP (aminoimidazole 4-carboxamide ribonucleotide). Two of the pockets were deemed less druggable, and by superimposing the binding sites (**Figure 19A**), it was clear that one of the less druggable binding sites had a conformational change in one residue (A60), which not only affected the shape of the superligand, but also the polar contact surface area (**Figure 19C, D**). The last remaining less druggable structure in this family shares a nearly identical conformation with the druggable instance, and it seems that minor differences affected the predicted outcome for these two cases, seeing as the top highest impacting descriptors as described by the SHAP values are similar (**Figure 19B, E**).

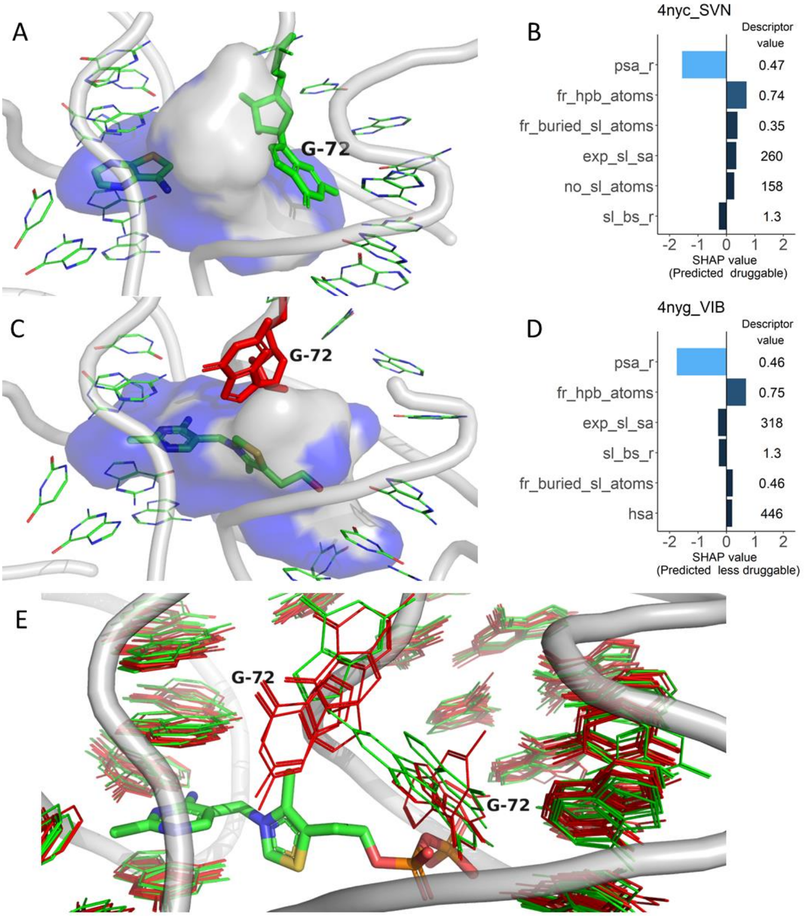

A
G-72
B
4nyc_SVN
Descriptor value
psa_r
fr_hpb_atoms
fr_buried_sl_atoms
exp_sl_sa
no_sl_atoms
sl_bs_r
0.47
0.74
0.35
260
158
1.3
-2 -1 0 1 2
SHAP value
(Predicted druggable)
C
G-72
D
4nyg_VIB
Descriptor value
psa_r
fr_hpb_atoms
exp_sl_sa
sl_bs_r
fr_buried_sl_atoms
hsa
0.46
0.75
318
1.3
0.46
446
-2 -1 0 1 2
SHAP value
(Predicted less druggable)
E
G-72
G-72

Figure 18: Druggability predictions for TPP riboswitch binding sites, with the flexible residue G72 highlighted. The surface of the superligand created by DrugPred_RNA as a negative print of the pocket is shown as a blob with the solvent exposed surface area colored gray and the remaining area colored blue. For the pockets shown in **(A)** and **(B)**, the individual SHAP values for the six most important descriptors are shown together with their descriptor values. The SHAP plots are labeled with the PDB IDs of the receptors and three-letter codes of the ligands found in each pocket (**B, D**). **(A)** TPP riboswitch binding site (PDB ID: 4nyc) in complex with a fragment screening hit (green sticks). **(C)** TPP riboswitch binding site (PDB ID: 4nyg) in complex with thiamine. **(E)** Superposition of all E. coli TPP riboswitch binding sites in the RNA-only set. Entries predicted to be druggable are colored green, and those predicted to be less druggable are colored red. For clarity, only the backbone (gray tube) from PDB ID 4nyc is shown.

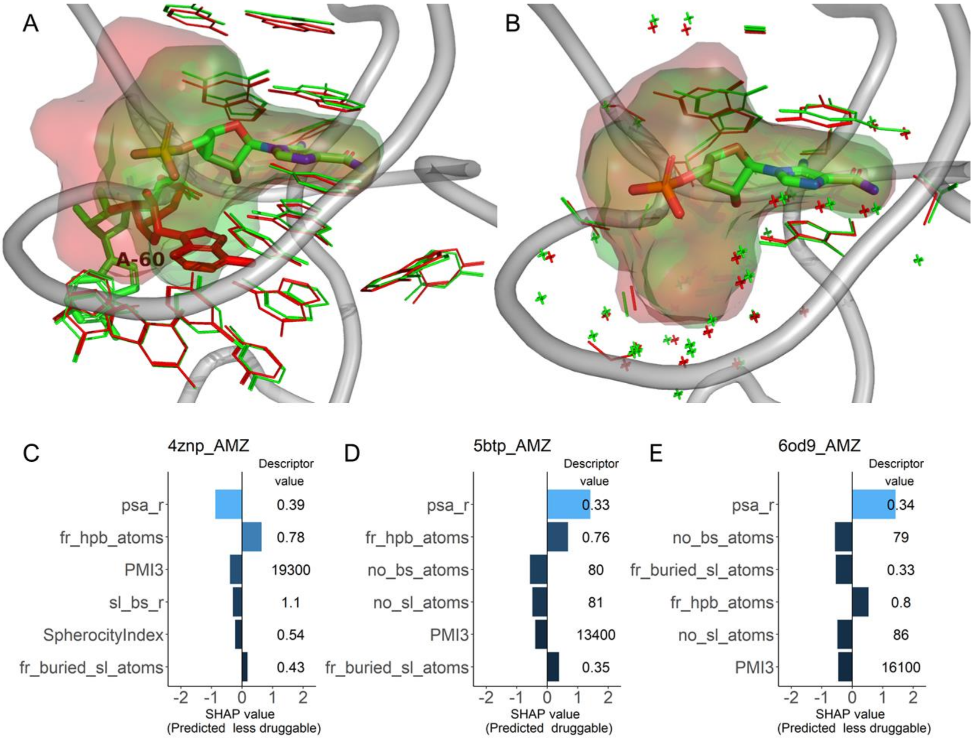


Figure 19: Superposition of the ZNP riboswitch binding sites bound to ZNP (thick sticks with green carbon atoms). The superligands created by DrugPred_RNA are shown as blobs. For clarity, only the backbone from 5btp is shown. **(A)** Superposition of the pockets of the structures with the PDB IDs 4znp (red, less druggable) and 5btp (green, druggable). The entire residues forming the binding sites are shown. The residue A60 is adopting two different conformations. **(B)** Superposition of the pockets of the structures with the PDB IDs 5btp (green, druggable) and 6od9 (red, less druggable). For clarity, only the atoms that DrugPred_RNA predicted to be in contact with the superligand are shown (thin sticks/crosses). **(C, D, E)** Individual SHAP values for the six most important descriptors for the displayed binding sites together with the descriptor values.

As has been shown, the binding site predictions provided by DrugPred_RNA are generally accurate and reliable. Some of the cases highlight the fact that conformational changes can impact the predictions of a binding site, but edge cases exist where only minute changes in binding sites can lead to different prediction outcomes. As such, it is advisable to use multiple crystal structures, if available, to assess the druggability of a single binding site.

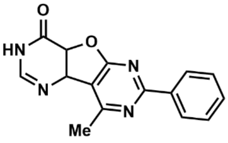
3.1.8 Finishing remarks

A new predictor based on a machine learning method which is compatible with both RNA and protein binding sites has been constructed and is able to provide predictions with high accuracy with comparable results to the previous version, DrugPred 2.0 when assessed on protein binding sites. When reviewing binding sites with expected druggable/less druggable properties (manual assignment of druggability), the predicted outcomes agreed with theory. All predictions for pockets predicted to be druggable which had a highly drug-like ligand with known affinity data, were correct, whereas in the five cases where a highly drug-like ligand was bound to a pocket predicted to be less druggable, four were likely incorrect. As such, this indicates that DrugPred_RNA has a higher false positive rate than a false negative rate, but the data set explored may be too small to draw this as a definite conclusion. DrugPred_RNA is sensitive to conformational changes, but this is not unique to neither this method or RNA binding sites.[71,149] The majority of cases provide high robustness scores (≥74%, depending on the dataset), which shows that DrugPred_RNA generally is quite accepting of small variations in binding site geometry.

As RNA binding sites are a less explored targets, DrugPred_RNA offers a wide array of potential new targets. The two combined sets explored 1078 binding sites, and while only a few (22) contained a highly drug-like ligand, 396 sites were predicted to be druggable. A significant amount of these are riboswitches, which underlines that these structures are interesting potential targets for new antibiotics. A high number of druggable pockets was also found in the ribosomal RNA, solidifying this as a good drug target and can shape efforts in the continuing development of antibiotics. As DrugPred_RNA was trained with descriptors suitable for both RNA and protein binding sites, the method can be applied on polymeric structures, or hybrid structures.

3.2 Design of New Ligands

The binding modes of FMN and ribocil A (**Figure 10**) suggest that the pyrimidinone moiety plays an important role for the affinities of these ligands and could serve as the basis for a new scaffold for FMN riboswitch ligands. Earlier findings from the Brenk group include **RSL-0035** (see **Figure 20**) as a hit compound from a virtual screening of potential ligands for the FMN riboswitch, which showed promising activity also in an SPR assay, but due to its scarcity, sufficient data could not be recorded. Among other modifications, this compound has a pyrimidinone moiety in the A ring (**Figure 11B, C**) instead of the pyrimidinedione ring found in FMN. Modelling of the binding mode using minimization also suggested a similar binding mode (**Figure 11D**) with the A, B and C rings sandwiched between residues A48/A85 and the pyrimidinone of the A ring base paired with A99 and forming an additional hydrogen bond with the backbone of A48.



RSL-0035

Figure 20: Structure of compound **RSL-0035.**

Based on the observed and postulated binding modes, a series of new ligands was designed, based on the following criteria:

1. Retain the interactions found for the ring system of FMN and postulated for **RSL-0035**, specifically the π/π-interactions with A85 and A48, base-pairing with A99 and hydrogen bonding with the backbone C2 hydroxyl of A48.
2. Explore possible new interactions within the binding site
3. Be synthetically feasible

A summary of all the proposed structure which were investigated in this thesis is shown in **Figure 21**.

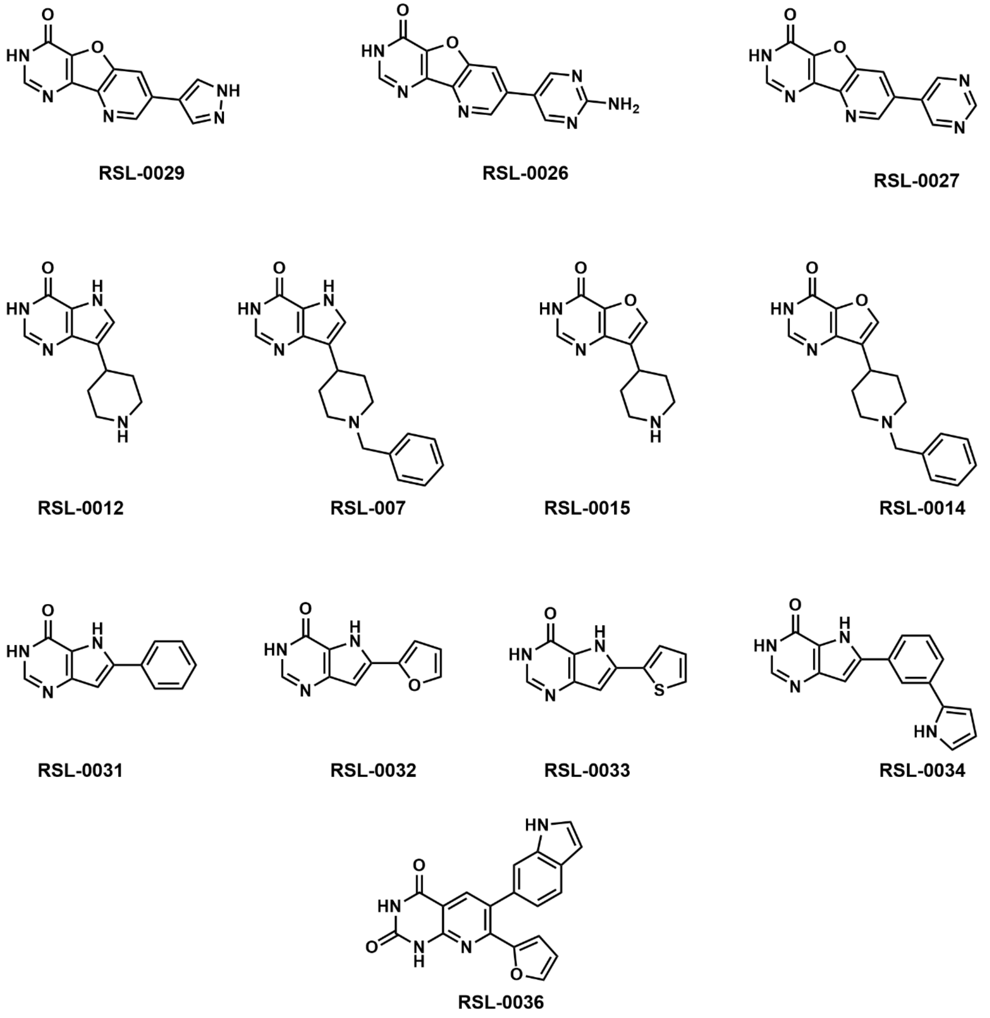

RSL-0029
RSL-0026
RSL-0027
RSL-0012
RSL-007
RSL-0015
RSL-0014
RSL-0031
RSL-0032
RSL-0033
RSL-0034
RSL-0036

Figure 21: Summary of all new proposed structures within this work.

3.2.1 Design and predicted binding modes of **RSL-0029, RSL-0026** and **RSL-0027**

Initial design of potential candidates was intended as derivatives of **RSL-0035**. In the modeled binding mode, only the central part participated in interactions with the binding site, therefore we assumed that the methyl group does not play a significant role. Further, it appeared that the phenyl ring did not provide any contribution to the proposed binding mode. Derivatives which shared the same central scaffold, minus the methyl group and with a variation with respect to the position of the aromatic substituent were therefore proposed as the next generation ligands. Based on these considerations, 58 suggested compounds were generated *in silico* and docked into the binding site.

For further evaluation, candidates were only chosen if their binding modes were proposed to be similar to those of FMN and **RSL-0035**. These structures were **RSL-0029, RSL-0026** and **RSL-0027** and their predicted binding modes are shown in **Figure 22.** While none of these compounds out-competed **RSL-0035** in terms of the energetic scores, the output from docking were quite similar to that what has been previously reported for this compound (Data not shown).[133]

Upon further minimization of the output poses of these riboswitch/ligand complex, we found that all compounds were able to participate in new interactions while retaining all the previously mentioned interactions, as shown in **Figure 22.** In the modelled binding mode, **RSL-0029** formed both a hydrogen bond with the C2 hydroxyl group of the ribose backbone in G84 and participated in an edge-face π/π-interaction with G62 (**Figure 22A),** which was a similar observation for **RSL-0027** (**Figure 22C)**. **RSL-0026** formed a hydrogen bond with the C2 hydroxyl group of the ribose moiety

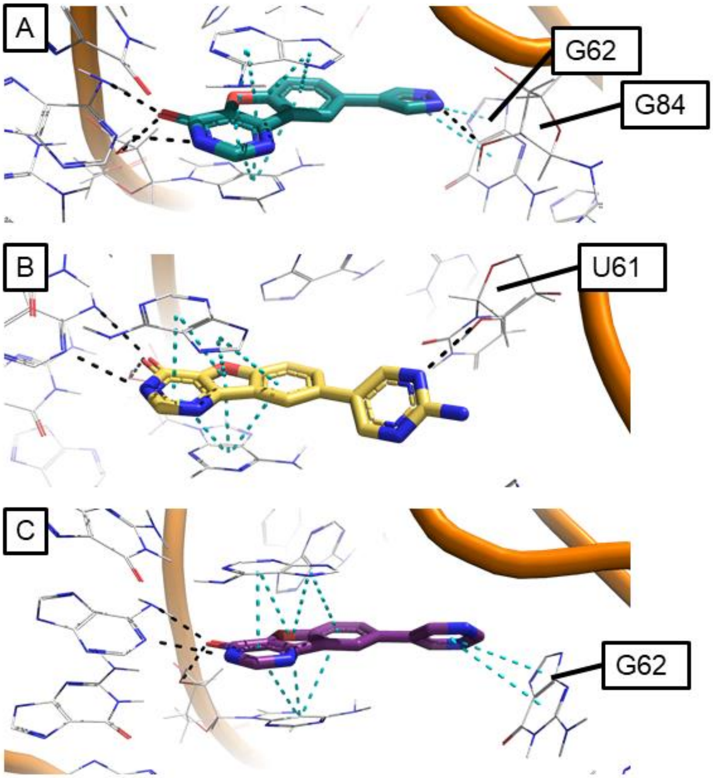


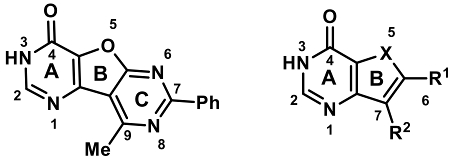
Figure 22: Minimized poses of the docking hits **A) RSL-0029** (green thick sticks), **B) RSL-0026** (yellow) and **C) RSL-0027** (magenta) complexed with the FMN riboswitch (grey thin sticks) after minimization. (PDB ID: 2yie, residues as grey thin lines, backbone as orange tubes. Putative hydrogen bonds drawn as black dashes, π /π-interactions drawn as cyan dashes.)

of U61, but the pose generated after minimization indicated that the aromatic substituent was not in a favorable position to form an interaction with G62 (**Figure 22B)**. Nevertheless, as this compound still had an overall desirable pose, it was still selected with the other two for further evaluation.

3.2.2 Design and predicted binding modes of **RSL-0012, RSL-0007, RSL-0015** and **RSL-0014**

Further design based on the three-ring scaffold was limited by the fact that branching out from the scaffold to attain any new π/π-interactions was very difficult without adding numerous connecting atoms, thus risking a severe loss of ligand efficiency. However, by removing the C ring of **RSL-0035,** a new central scaffold which offered the opportunity to branch out in new vectors was identified (**Figure 23**).

After molecular minimization was applied on structures based upon this scaffold, a similar binding mode as seen for FMN and predicted for **RSL-0035** was found with sandwiching π/π-interactions between A85 and A48, base-pairing with A99 and hydrogen bonding with the C2 hydroxyl group in A48. The binding modes of derivatives are described later within this chapter.



RSL-0035

X = O, NH

Figure 23: Numbering of atoms and labeling of the rings in **RSL-0035** and bicyclic analogues.

At this point, it was decided to step away from using docking and molecular minimization as design tools and to only use molecular minimization instead. One obvious advantage with molecular minimization over docking is the fact that some flexibility of the binding site can be taken into account.

However, a drawback of only employing molecular minimization as a modelling method is that there is no guarantee that the proposed conformation truly represents a global energetical minimum within the binding site. However, it would be unlikely that a similar compound could gain a larger hydrophobic contact surface area than what is already proposed without losing the hydrogen bonds between the A ring and the binding site, and as so it is unlikely an alternate pose would represent a true global energetical minimal pose. It is also important to keep in mind that a molecular minimization model provides only a snapshot of one energetic minimum of the binding site/ligand-ensemble, of which there might be several. As so, only a crystal structures of the proposed ligands bound to the FMN riboswitch could provide insight into what the true binding mode of the ligands would be.

Several structures with substituents in the C7 position of the new scaffold were explored (**RSL-0012**, **RSL-0007**, **RSL-0015**, **RSL-0014**), and the suggested binding modes are shown in **Figure 24**. After identifying a nitrogen atom in the 5-position as a potential hydrogen bond donor to the endocyclic oxygen in the A49 ribose backbone, analogues with both a nitrogen (**RSL-0012**, **Figure 24A, RSL-0007, Figure 24C**) and oxygen (**RSL-0015**, **Figure 24B, RSL-0014, Figure 24D**) atom in this position were

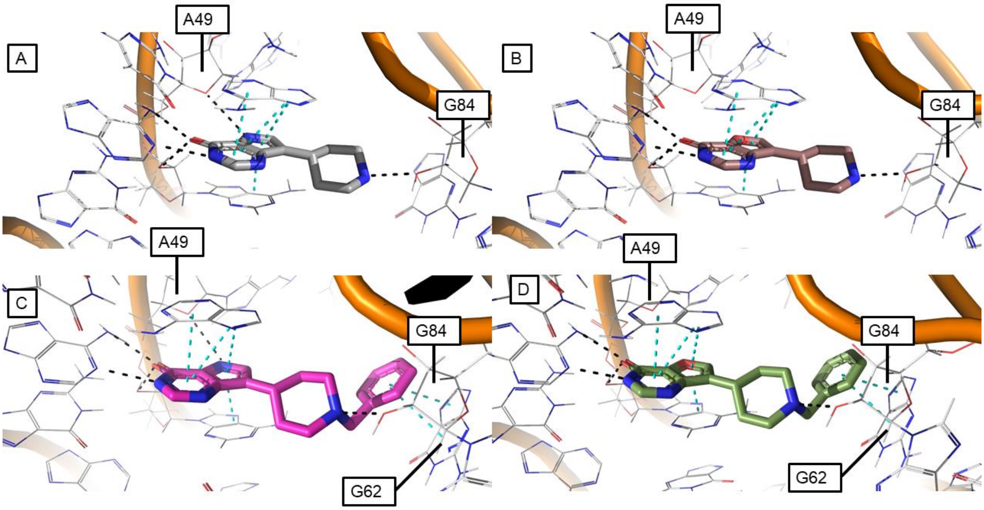


Figure 24: Minimized poses of **A) RSL-0012 (**Grey thick lines), **B) RSL-0015** (brown thick lines), **C) RSL-0007** (magenta thick lines, **D) RSL-0014** (green thick lines) in the FMN riboswitch. (PDB ID: 2yie, residues as grey thin lines, backbone as orange tubes. Putative hydrogen bonds drawn as black dashes, π /π-interactions drawn as cyan dashes.)

proposed. Adding a piperidyl substituent to the C7-position should allow for formation of a hydrogen bond with the C2 hydroxyl group in G84, while a benzyl group on the piperidyl *N*-atom can potentially establish a very beneficial π/π-stacking interaction with G82 (**Figure 24C, D**). The geometry of this π/π-interaction differed from what was found for previous examples, where here the substituent stacked on the residue instead of forming an edge-face interaction.

Comparing the predicted binding mode of for example **RSL-0007** with ribocil A (**Figure 25A, B**), it revealed that the benzylic substituent in the former could adopt a similar stacking interaction with G62 as ribocil A. Furthermore, the two tertiary amines in the structure did not over, and it seemed that **RSL-0007** and **RSL-0014** were in a geometrically more favorable position to form a hydrogen bond with the backbone of G84 compared to ribocil A. The central scaffold which was supposed to form hydrogen

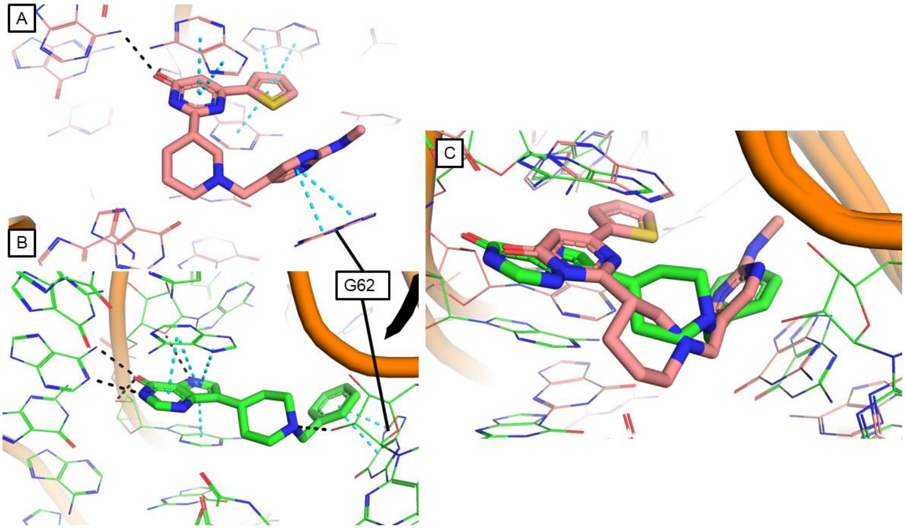


Figure 25: A) Crystal structure of ribocil A (thick lines) in the FMN riboswitch (thin lines) **B)** Minimized pose of **RSL-0007** (thick lines) in the FMN riboswitch (thin lines). **C)** **RSL-0007** and ribocil A superimposed. (PDB ID: 2yie, residues as grey thin lines, backbone as orange tubes. Putative hydrogen bonds drawn as black dashes, π /π-interactions drawn as cyan dashes.)

bonds between ring A and A99 are also slightly tilted (**Figure 25C**) when compared to the pyrimidinone ring of ribocil A.

Apart from being a potential hydrogen-bond donor, the piperidyl substituent should also provide flexibility in the side chain to more easily allow for interactions with the binding site to be formed. It also provided a site of protonation, which is beneficial for solubility. Exploring the two-variable matrix of hydrogen-bond and stacking interactions would allow for an establishment of SAR. A benzylic substituent on a piperidine nitrogen atom opens for preparation of a number of possible analogues, as a wide range of aromatic aldehydes (a typical building block for the synthesis of these compounds) are commercially available.

3.2.3 Design and predicted binding modes of **RSL-0031, RSL-0032**, **RSL-0033** and **RSL-0034**

The next series of compounds were designed with the assumption that having a nitrogen atom within the B ring of the bicyclic scaffold (**Figure 23**) would be beneficial as it could form a hydrogen bond with A49, and we were interested in exploring substituents in the C6-position. Thus **RSL-0031, RSL-0032, RSL-0033** and **RSL-0034** were proposed, and the binding poses after minimization are shown in **Figure 26.**

A phenyl group (**RSL-0031, Figure 26A**) in the C6-position could form a π/π-interaction with the nucleobase of A49. However, the pose obtained after molecular minimization, suggested that the angle and distance between these aromatic moieties was not ideal. The synthetic route to **RSL-0031** allowed other aromatic substituents to be explored, and as such, **RSL-0032** and **RSL-0033** were proposed. Here, the binding poses obtained after molecular minimization had a more optimal π/π-stacking interaction between A49 and the furanyl and thiofuranyl substituents (**Figure 26B, C**). Molecular minimization suggested that **RSL-0034** formed an edge-face interaction with G62 (**Figure 26CD**, similar to what has previously been seen with **RSL-0029**, **RSL-0026**, (**Figure 22A, C**), and **RSL-0007** (**Figure 26C**), and also a hydrogen bond with the endocyclic ribose oxygen of A85, not observed previously.

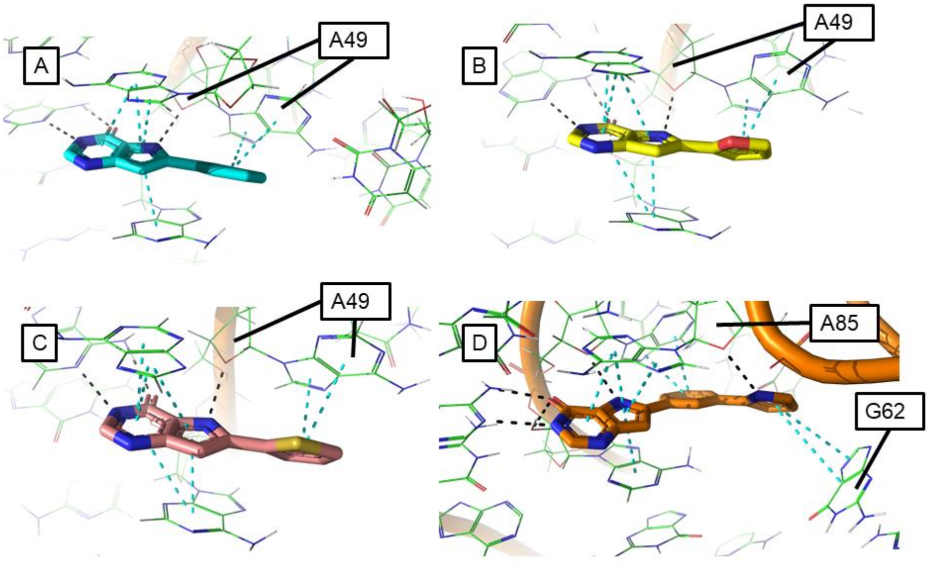


Figure 26: Minimized poses of **A) RSL-0031** (thick lines), **B) RSL-0032** (magenta thick lines), **C) RSL-0033** (brown thick lines) and **D) RSL-0034** (green thick lines) in the FMN riboswitch. (PDB ID: 2yie, residues as grey thin lines, backbone as orange tubes. Putative hydrogen bonds drawn as black dashes, π /π-interactions drawn as cyan dashes.)

3.2.4 Design and predicted binding mode of **RSL-0036**

The last compound, **RSL-0036,** explored in this study had a completely different scaffold than previous structures, and was inspired by the work of Lindeman *et al.*[150], which was noticed in a literature search in a different project within the Haug group. This compound, which contains a pyrimidinedione ring fused with a pyridine ring, has two aromatic substituents which are attached to the center scaffold in similar vectors as both R1 and R2 of the bicycle scaffold previously described **(Figure 23).** The proposed binding mode of the compound in the FMN riboswitch binding site obtained through molecular minimization is shown in **Figure 27.** Again, this binding mode was similar to the previously discussed compounds and fulfilled the minimum criteria we were interested in: the fused rings of the central scaffold were sandwiched between A85/A48, the pyrimidinedione moiety base paired with A99 and the additional hydrogen bonding was observed to the backbone of A48. Furthermore, the furanyl ring was suggested to form a π/π stacking interactions with A49 (similar to **RSL-0032**) and the indolyl substituent could form a π/π edge/face-interaction with G62. Hydrogen bonding between the indolyl substituent and the C2 backbone of G84 (similar to **RSL-0012** to **RSL-0014)** was an additional possibility.

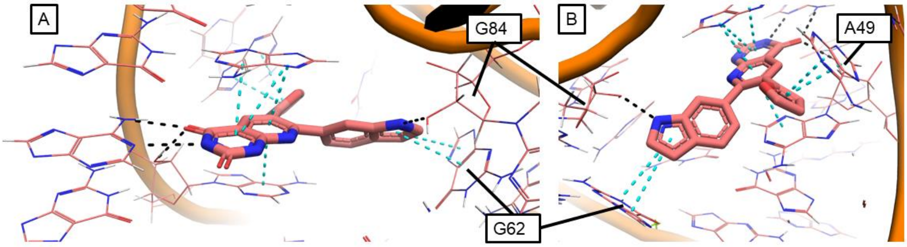


Figure 27: Alternate views (**A**, **B**) of the minimized poses of **RSL-0036** (orange thick sticks), in the FMN riboswitch. (PDB ID: 2yie, residues as grey thin sticks, backbone as orange tubes. Putative hydrogen bonds drawn as black dashes, π /π-interactions drawn as cyan dashes.)

3.2.5 Final remarks regarding *in silico*-designed compounds

Collectively, the proposed compounds were designed with the intent of maintaining proposed key interactions with the FMN riboswitch while still exploring structural diversity. As can be seen in **Table 7**, the compounds are fragment-like, with MW < 300, close to 3 hydrogen bond donors and acceptors and clogP <3. Furthermore, the QED-values implied that the compounds are drug-like, with most scores close to or higher than the mean QED score of approved drugs (0.539).[151]

Table 7: Molecular descriptors of proposed structures, calculated with RDKit. N_{HBD} = Number of hydrogen bond donors, N_{HBA} = Number of hydrogen bond

Compound	MW	N_{HBD}	N_{HBA}	clogP	QED score
RSL-0035	278.27	1	5	2.43	0.58
RSL-0029	253.22	2	5	1.45	0.53
RSL-0026*	281.26	2	6	0.39	0.51
RSL-0027	265.23	1	6	1.52	0.56
RSL-0012*	219.27	3	2	-0.308	0.62
RSL-0007*	309.39	3	2	1.21	0.68
RSL-0015*	220.25	2	3	-0.0431	0.71
RSL-0014*	310.38	2	3	1.48	0.77
RSL-0031	211.22	2	2	1.92	0.65
RSL-0032	201.19	2	3	1.51	0.63
RSL-0033	217.25	2	3	1.98	0.65
RSL-0034	276.30	3	2	2.91	0.53
RSL-0036	344.33	3	4	3.02	0.46

* Protonated states

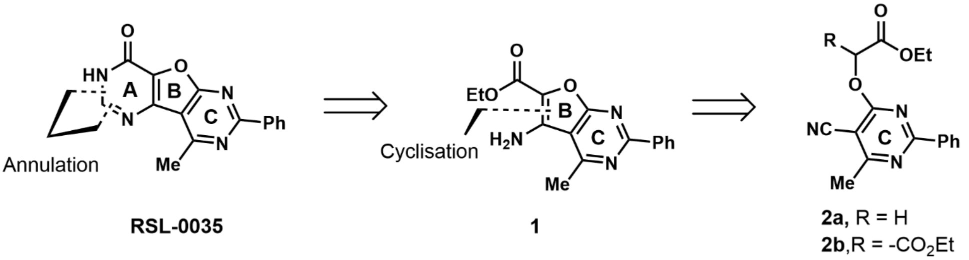
The synthesis of **RSL-0035, RSL-0031, RSL-0032, RSL-0033, RSL-0034 and RSL-0036** is described in the following chapter. The synthesis of **RSL-0029, RSL-0026** and **RSL-0027** has been described in the master thesis of Valeriia Burova.[152] The synthesis of **RSL-0015, RSL-007** and **RSL-0014** is described in the master theses of Heidi Kristine Vintermyr and Andrea Osvoll Årdal.[153,154] The basis for the syntheses carried out by Burova, Vintermyr and Årdal was worked out by Dr. Muhammed Zeeshan, who also prepared **RSL-0012.**[155]

3.3 Synthesis of proposed FMN ligand candidates

3.3.1 Synthesis of hit compound **RSL-0035**

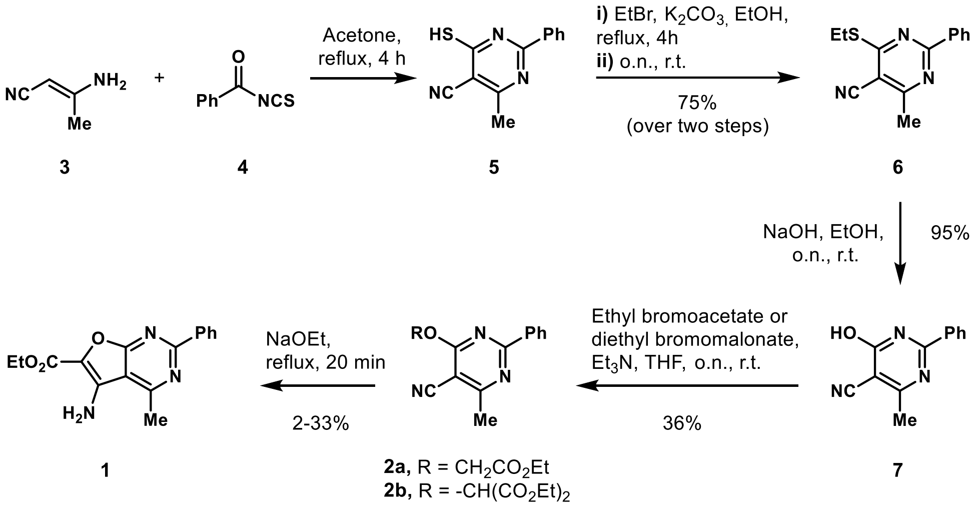
RSL-0035 was originally purchased from Specs, but at the onset of this work it was no longer commercially available**.** To confirm the affinity of this proposed hit compound to the FMN riboswitch, in-house synthesis was initated to gain more material, with the added goal of verifying that the structure of the originally purchased compound was correct. The synthesis of RSL-0035 has been previously reported by Moneam *et al.*[156] The strategy of this synthesis is outlined in

Scheme 1. The B and A rings were gradually built from ring C in a cyclisation and subsequent annulation reaction.


O
HN
A
B
C
N
N
Ph
Me
Annulation
RSL-0035
EtO
O
O
B
C
N
N
Ph
H2N
Me
Cyclisation
1
R
O
OEt
O
NC
N
C
N
Ph
Me
2a, R = H
2b,R = -CO2Et

Scheme 1: Retrosynthetic analysis of **RSL-0035.**

We followed essentially the same strategy in our synthesis of **RSL-0035**, and the initial steps are outlined in **Scheme 2.**

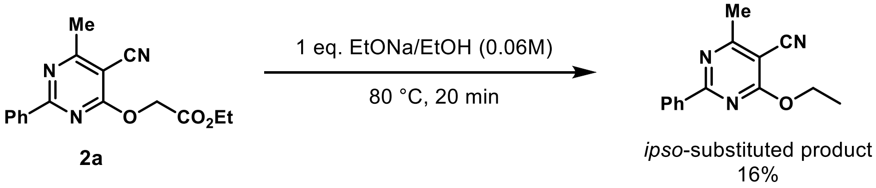

Acetone, reflux, 4 h
i) EtBr, K_2CO_3, EtOH, reflux, 4h
ii) o.n., r.t.
75% (over two steps)
NaOH, EtOH, o.n., r.t.
95%
Ethyl bromoacetate or diethyl bromomalonate, Et_3N, THF, o.n., r.t.
36%
NaOEt, reflux, 20 min
2-33%
3
4
5
6
7
1
2a, R = CH_2CO_2Et
2b, R = $-CH(CO_2Et)_2$

Scheme 2: Synthesis of aminoester 7, a precursor to **RSL-0035**.

Pyrimidine **5** was prepared first by refluxing 3-aminocrotonitrile (**3**) and isothiocyanide **4** in acetone, however initial attempts at purification of the crude product resulted in low yields.[157] This step was instead telescoped with the alkylation step to provide **6** in good yield. The subsequent hydrolysis of **6** to remove the thioethyl group furnished phenol **7**, which was then alkylated using either ethyl bromoacetate or diethyl bromomalonate. Purification of **2a** or **2b** using silica flash chromatography provided relatively low yields (25-40%), and the crude products were used instead in the subsequent step. Cyclisation of esters **2a**/**2b** to aminoester **1** follows a Thorpe-Ziegler-like mechanism,[158,159] with initial deprotonation of the ester alpha position and subsequent intramolecular nucleophilic attack on the nitrile to form the five-membered ring. The subsequent step is dependent on the ester moiety used in the reaction; if an ethyl ester is used, the cyanocarbon is attacked by the deprotonated alpha ester before aromatization, if the ester is a malonate, loss of a carbon dioxide molecule transpires before aromatizaiton. Following the original reported conditions from Moneam *et al.*, freshly prepared sodium ethoxide from sodium and absolute ethanol was used as a base for the cyclization reaction with the malonate **2b**. These conditions provided aminoester **1** in only 2-10% yield after purification as opposed to the 83% originally reported. It is possible presence of water in the ethanol might have contributed to the low yields in this reaction.

Although the original paper explored cyclization both with the mono- and diester **2a** and **2b**, no details were provided regarding the difference in outcome for the reactions other than that the latter outperformed the former. There is also ample evidence in the literature that such a cyclization is as viable with a monoester starting material as with a malonate; Sleebs *et al.*[160] reported that a multitude of benzothiophenes and benzofurans could be formed through reactions with ethyl bromoacetate and bromoacetamide using potassium carbonate, while Keshk *et al.*[161] reports the same on a similar furo-[2,3]-pyridine.

Different conditions were therefore investigated to increase the yield of the cycylisation, as summarized in **Table 8**. The conditions from Moneam *et al.,* using ethoxide (Entries 1-4) were not very productive. Use of stoichiometric amounts of base led to the isolation of an *ipso*-substituted product (**Scheme 3**), a product which draws some similarities with the findings of Yamanaka *et al,* where it was found that alkyloxy groups could be substituted by nucleophilic reagents.



Scheme 3: Side-product isolated using 1 equivalent of sodium ethoxide.

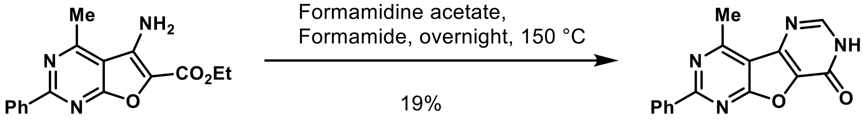
The conditions explored by Sleebs *et al.*, using potassium carbonate or cesium carbonate (entries 5-10) proved futile. Strong, non-nucleophilic organic bases were also explored (entries 11-13). DBU proved to be ineffective, while potassium *tert*-butoxide in combination with the monoethyl ester **2a** only gave a mixture of the hydrolyzed carboxylic acid derivative of **1** and starting material**.** While high-quality absolute ethanol was used for these reactions, it is possible residual water in the solvent is the reason for this hydrolysis.

Cyclization using potassium *tert*-butoxide of the malonate ester (entry 13) proved to be a quick reaction, requiring only 10 minutes at room temperature before full conversion was observed *via* TLC analysis, giving aminoester **7** in 50% crude yield. This provided sufficient starting material for the final step in the synthesis of **RSL-0035.**

Table 8: Conditions explored for enhancing the yield in the cyclization reaction from 2a/2b to aminoester 1. n.r = no reaction, r.t. = room temperature, o.n. = overnight, m = min.

Entry	**Electro-phile**	**T [°C]**	**Time**	**Base**	**Solvent**	**Comment**	**Result**
1	**2a**	80	20 m	0.3 eq. EtONa	EtOH	Fresh base	10%
2	**2a**	80	20 m	0.3 eq. EtONa	EtOH	Not fresh	Decomposed
3	**2a**	80	20 m	1.0 eq. EtONa	EtOH	-	ipso-substitution
4	**2b**	80	20 m	0.3 eq. EtONa	EtOH	-	33%
5	**2b**	80	o.n.	2 eq. K_2CO_3	DMF	-	1.6%
6	**2b**	80	4 days	2 eq. K_2CO_3	DMF/EtOH	-	5.6%
7	**2a**	80	2 days	2 eq. Cs_2CO_3	DMF	-	10% + Hydrolysis
8	**2b**	80	20 m	2 eq. Cs_2CO_3	DMF	w/ mol. sieves	NR
9	**2a**	80	20 m	2 eq. Cs_2CO_3	DMF	w/ mol. sieves	2.6
10	**2b**	r.t.	20 m	2 eq. Cs_2CO_3	DMF	-	10%
11	**2b**	80	o.n.	DBU	DMF	-	n.r.
12	**2a**	80	1 h	1.0 eq. KOtBu	EtOH	-	Hydrolysis
13	**2b**	r.t.	10 m	1.0 eq. KOtBu	EtOH	-	~50%*

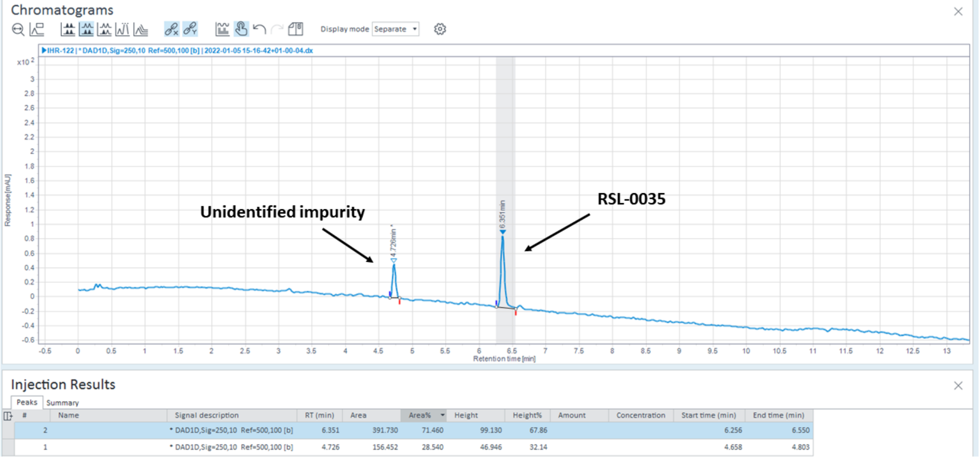
* Crude yield



1 **RSL-0035**

Scheme 4: Annulation of aminoester **1** with formamide/formamidine to yield the final product **RSL-0035.**

The final annulation reaction between aminoester **1** and formamide/formamidine acetate to form ring A in **RSL-0035 (Scheme 4)** was succesfully carried out after a necessary adjustment of reaction temperature. The original procedure called for the reaction to be refluxed in formamidine (b.p 210 °C), while formamide is known to decompose at temperatures >185 °C into CO, CO_2 and HCN.[162,163] Annulations of similar aminoesters with formamide to give pyrimidinones has previously been reported to be effective at 150 °C.[160] An attempt was made to see if the reaction would yield product at lower temperatures, avoiding unnecessary thermal decomposition of product and solvent and avoiding excessive energy use (The 6th and 10th principle of green chemistry).[164] After overnight stirring at 150 °C, **RSL-0035** was obtained after purifying the crude product using reverse phase chromatography on an autoflash instrument followed by purification using reverse phase HPLC. The purification was complicated by the fact that **RSL-0035** seems to have limited solubility in DMSO (around 10-20 mg/mL), which made purifying larger quantities very time consuming. Furthermore, HPLC analysis of **RSL-0035** after several months of storage at 5 °C showed decomposition (**Figure 28**) making it necessary to prepare this compound immediately prior to submission for activity evaluation. Attempts at identifying the unknown impurity were undertaken using LC-MS, but this failed to provide any mass beyond that which was expected of **RSL-0035**.

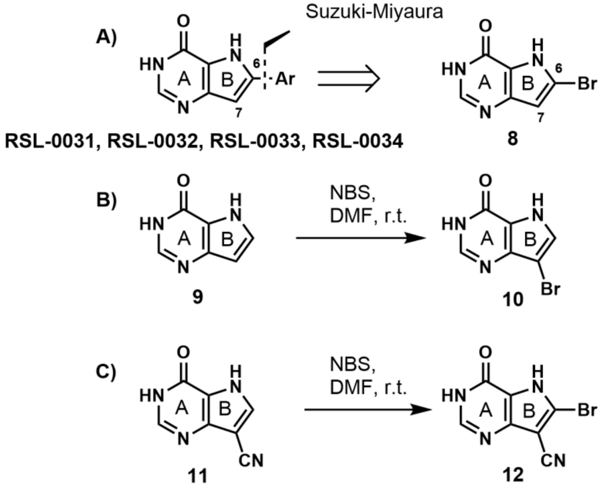

Chromatograms
Display mode
Separate
▶IHR-122 | * DAD1D,Sig=250,10 Ref=500,100 [b] | 2022-01-05 15-16-42+01-00-04.dx
x10 2
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Response [mAU]
Unidentified impurity
RSL-0035
4.726min *
6.351min
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5
9
9.5
10
10.5
11
11.5
12
12.5
13
Retention time [min]
Injection Results
Peaks
Summary

#	Name	Signal description	RT (min)	Area	Area%	Height	Height%	Amount	Concentration	Start time (min)	End time (min)
2		* DAD1D,Sig=250,10 Ref=500,100 [b]	6.351	391.730	71.460	99.130	67.86			6.256	6.550
1		* DAD1D,Sig=250,10 Ref=500,100 [b]	4.726	156.452	28.540	46.946	32.14			4.658	4.803

Figure 28: Analysis of **RSL-0035** stored in DMSO for ~3 months.

3.3.2 Synthesis of in silico designed compounds **RSL-0031, RSL-0032, RSL-0033** and **RSL-0034**

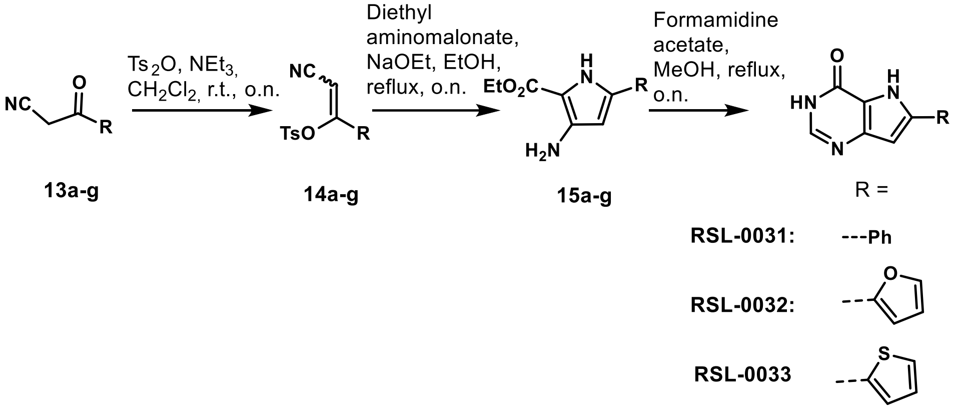
The π/π-interactions between the aromatic substituent in the C6-position of the bicyclic scaffold in Scheme 5 and A49 which was described in the previous chapter would make it ideal to find a synthetic route which allowed for a facile route to multiple analogues with different aromatic substituents in this position. An ideal starting material would be the 6-substituted bromide **8** (**Scheme 5A**), which could act as an electrophilic coupling partner in a Suzuki-Miyaura coupling. However, none of the suppliers we investigated were able to supply this chemical in multi-gram scale. A literature search revealed that the reaction of bicycle **9** with NBS or bromine leads to bromination at the 7-position to give bromide **10 (Scheme 5B)**.[165,166] We found no other examples where other bromination agents were employed, but we assumed the electronic properties of **9** would always favor halogenation in the 6-position. The closest analogue example provided in the literature was the bromination of a 7-cyanosubstituted analogue **11** to give bromide **12** (**Scheme 5C**) found in two separate patents, however, considering that this starting material notably differs from what was modeled in the last chapter and the chemical is also not commercially available in any significant quantities further complicated this as a viable strategy.[167,168]



Scheme 5: A) Potential synthetic strategy to obtain **RSL-0031, RSL-0032** and **RSL-0033** B) Bromination of the central scaffold leads to undesired products[165,166] C) Potential alternative to obtain desired bromination.[167,168]

A method reported by Chen *et al.* has described the synthesis of **RSL-0031** in a 3-step synthesis from benzoyl acetonitrile.[169] The obvious drawback of this strategy is no possibility of late-stage functionalization, as the aromatic moiety is introduced in the first step of the synthesis. If **RSL-0031** or any of its analogues would prove to be active in a biological assay, focus on a more flexible synthetic route could become relevant, but meanwhile the short synthetic route remained the best option to obtain material which could serve as a proof of concept.

We were able to follow the reported synthesis of **RSL-0031** by Chen *et al.* using benzoyl acetonitrile **13a** as outlined in **Scheme 6**. After initial tosylation of ketone **13a** with *para*-toluenesulfonic anhydride, the intermediate tosylate (**14a**) was isolated in a 58% yield. In our case, the *E*:*Z* isomer ratio, as determined by NMR, was found to be 1:6 (*vs.* 1:3 reported earlier). After annulation of **14a** with diethyl aminomalonate and sodium ethoxide as a base, amino ester **15a** was obtained in a better yield than reported (71% *vs.* 47%). Subsequent annulation using formamidine acetate in refluxing methanol gave the target compound **RSL-0031** in sufficient purity without the need for any purification beyond simple filtration. The yield of this compound (37%) is comparative to the previously reported yield (46%).

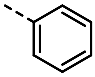
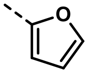
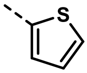
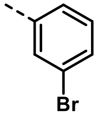
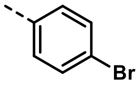
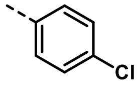
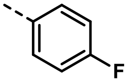

Ts$_2$O, NEt$_3$, CH$_2$Cl$_2$, r.t., o.n.
Diethyl aminomalonate, NaOEt, EtOH, reflux, o.n.
Formamidine acetate, MeOH, reflux, o.n.
NC
O
R
TsO
EtO$_2$C
H$_2$N
HN
N
H
13a-g
14a-g
15a-g
R =
RSL-0031:
---Ph
RSL-0032:
RSL-0033
O
S

Scheme 6*:* Synthetic approach to **RSL-0031** and analogues. A full description of the substituents is listed in **Table 9.**

A number of other 3-aryl-3-oxopropane nitriles **13b-g** were also subjected to the same synthetic protocol as shown in **Scheme 6**. It is speculated that only the less thermodynamically stable Z-diastereomers of **14a-e** can participate in the following Thorpe-Ziegler reaction to give the pyrroles **15a-e**. The high temperature in this transformation should enable isomerization of the unreactive *E*-diastereomer into the reactive *Z*-diastereomer (**Scheme 7**), and as such, for compounds **14b-g,** the diastereomers were not separated and the crude product following tosylation was used in the next steps. ^{1}H-NMR showed clear conversion into tosylates **14b-g** (not shown) and the overall mass balance proved to be excellent.

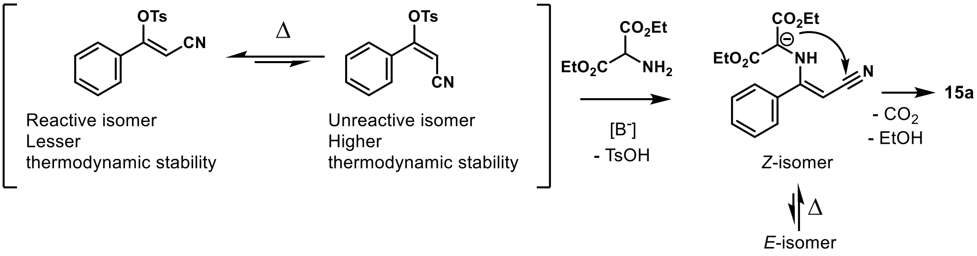
For the subsequent annulation reactions, yields for the 3-amino-4-ethoxycarbonyl substituted pyrroles **15b-g** varied from mediocre to abysmal (**Table 9**), where in some cases we were not able to isolate any product which could be structurally elucidated. While TLC analysis indicated complete consumption of starting material, purification of these crude products was unsuccessful. ^{1}H-NMR of the crude product revealed that a complex mixture was formed but did not offer any clear insight into which compound was the major product **(Figure 29A).**

Table 9: Isolated yields of compounds **15a-g.**

Compound name	**R =**	**Isolated yield**
15a		71%
15b		45%
15c		24%
15d		-*
15e		15%**
15f		-*
15g		-*

* Unable to isolate pure product; only trace material recovered

** ^{1}H-NMR matches with published data, unable to confirm with ^{13}C-NMR and HRMS


OTs
CN
Δ
Reactive isomer
Lesser
thermodynamic stability
Unreactive isomer
Higher
thermodynamic stability
CO2Et
EtO2C
NH2
[B-]
- TsOH
NH
N
Z-isomer
Δ
E-isomer
15a
- CO2
- EtOH

Scheme 7: Mechanism of Thorpe-Ziegler reaction.

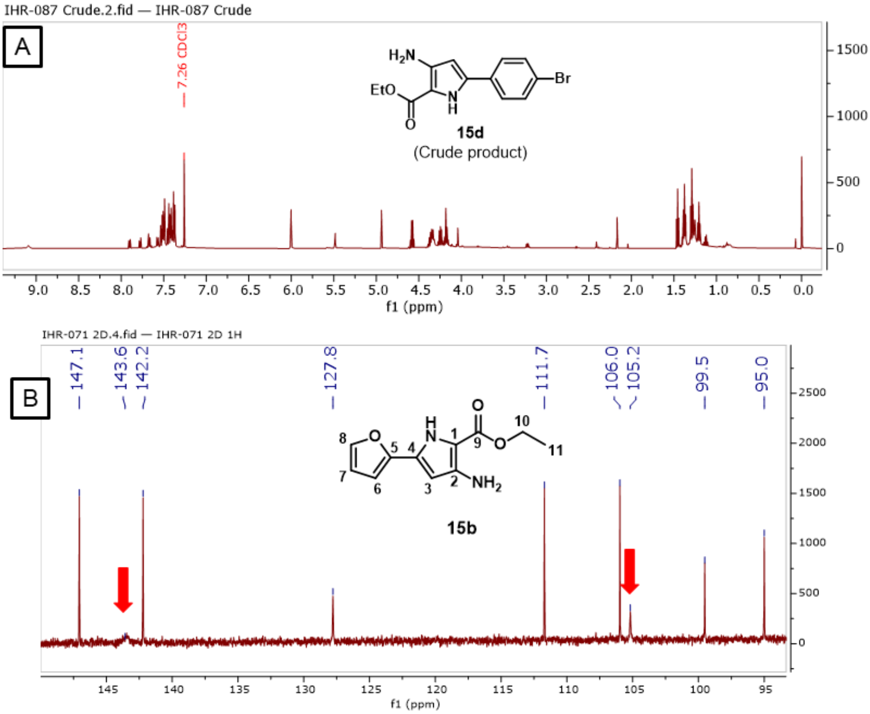


Figure 29: A) ^{1}H-NMR of the crude reaction mixture of 1**5d B)** ^{13}C-NMR of compound **15b**. Highlighted signals are broad signals from carbon 1 (δ = 105.7) and carbon 2 (δ = 143.6).

Another peculiarity of these compounds is the difficulty in obtaining ^{13}C-NMR signals for the quaternary C-atoms. For example, two of the ^{13}C-signals in **15b** are extraordinarily broad and low in intensity (**Figure 29B**), with one signal ($\delta = 143.6$) so broad that is almost indistinguishable from noise. HSQC showed no direct correlation with an attached proton and as such we assume these signals to correlate with quaternary carbon atoms (C1 and C2, **Figure 29B)**. Experiments with longer relaxation times (60 s) and inverse gated decoupling was performed to see if the intensity of these signals could be enhanced, but this proved to be unsuccessful. We were also not able to obtain two-dimensional spectra of sufficient quality to assess the relationship of this signal. The question then arises if these signals truly stem from compound **15b** or are due to impurities. One possible explanation of signal broadening could be paramagnetic relaxation caused by complexation with certain metal impurities. However, as this would probably also affect other carbon signals in the vicinity (such as C3, C4 and C9) and the fact that our syntheses does not involve any metals such as Cu^{2+}, Mn^{2+} or similar, we rule this out as a likely explanation. It is noteworthy that the report from Chen *et al.* does not include ^{13}C-NMR data for the intermediate **15a** or most of the other pyrrole amino diethyl ester analogues synthesized in their work.[169] Furthermore, a recent synthesis of **15e** reports only ^{1}H and high resolution MS data, but not ^{13}C-NMR data.[170]

In our hands, we were not able to confirm the product **15e** we isolated from the reaction mixture due to missing ^{13}C-NMR signals and failure to obtain HRMS data. However, ^{1}H-NMR data matches with the published data.[170] As shown in subsequent paragraphs, the use of **15b** successfully gave the final compound **RSL-0032**, which also confirms that this compound indeed is formed in the reaction from **14b.** We therefore turned to predictive methods to see if the broad carbon signals for C1 and C2 in **Figure 29B** were close to what would be expected (**Table 10**). While predicted shifts of the ^{13}C-NMR shifts based on calculations from various empirical methods were close for many of the experimental signals from **15b**, only one method (ChemDraw) came close to predicting the shift for C2, and none came close to the value for C1. The disparity of the predictions for the pyrrole carbons between methods is also noteworthy. It is possible these empirical methods fail because they do not take into account a more complex molecular behavior of **15b.** As such, a low-cost quantum mechanical (QM) computation was applied after minimization of **15b** to see if more insight could be gained. As can be clearly seen in **Table 10**, the predictions from NWChem clearly gave the closest predictions to the experimental data. Exchange processes and tautomerism are known to affect signal broadening in NMR, and as such a possible explanation is formation of an intramolecular hydrogen bond between the amino group attached to carbon 2 and either of the oxygens in the ester moiety, which would further complicate the electronic behavior of the pyrrole ring in **15b**. These results give insight into why ^{13}C-NMR shifts seem to be missing in the literature for such compounds, but more work needs to be done to be elucidate these observations. These findings, in combination with challenges faced in the Thorpe-Ziegler annulation of **6a** and **6b,** shows that this transformation is challenging.

Table 10: Experimental and predicted ^{13}C-NMR shifts for compound 15b. C# = Carbon number (as shown in **Figure 29B**). δ_E = Experimental shitfs, MAE = mean absolute error. Predicted carbon shifts for NWChem are calculated for the minimum energetic conformation of 10b. Closest predicted value for each shift to experimental values are highlighted with an asterisk.

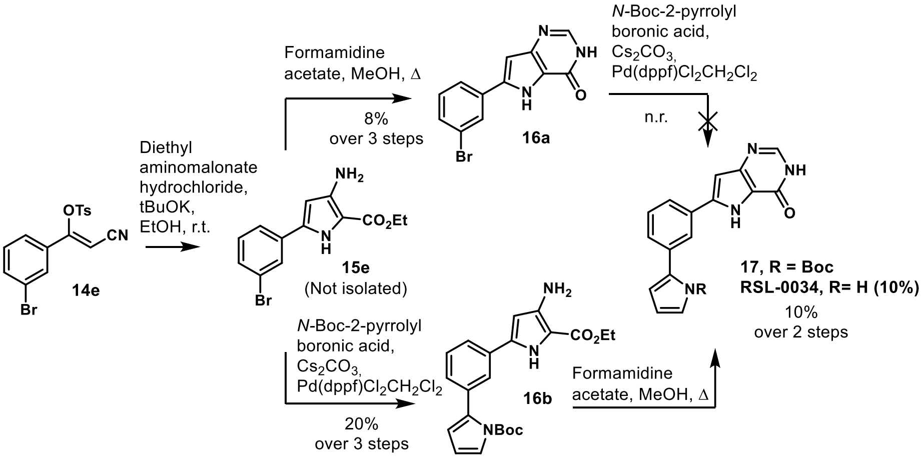
C#	δ_E	ChemDraw	NMRDb.org[171]	NMRShiftDB[172]	NWChem* [173]
C1	105.2	120.1	121.4	136.4	106.9*
C2	143.6	142.9*	132.5	121.9	146.8
C3	95.0	102.3	104.2	103.4	92.2*
C4	127.8	133.6	133.7	136.4	128.3*
C5	147.1	157.7	143.3	152.4	147.6*
C6	106.0	107.1	110.4	115.0	106.5*
C7	111.7	112.0	112.0	111.7*	111.1
C8	142.2	142.9	143.6	144.7	142.0*
C9	161.1	160.1	161.7*	160.0	161.3
C10	58.5	60.9	61.2	60.3*	60.8
C11	14.7	14.1	14.2	14.2	14.4*
MAE	-	4.12	5.1	8.2	1.2

** Calculations done by Associate Professor Nils Åge Frøystein*

Conversion of amino esters **15b** and **15c** to **RSL-0032** and **RSL-0033**, respectively were completed by reaction with formamidine acetate in refluxing methanol and provided an acceptable yield (**Table 9**). **RSL-0032** and **RSL-0033** were purified using reversed-phase (RP) flash chromatography on a C18 column (denoted C18 column onwards) and **RSL-0033** was further purified using semi-prep HPLC on a C18 column.

RSL-0034 was prepared from amino ester **15e**, however, attempts at isolating this compound proved futile (**Table 9**). The conditions which were used to yield the fused furane **7** (Entry 13, Table 8) were applied (as they were found to be effective in the Thorpe-Ziegler reaction, **Scheme 8**). The following annulation with formamidine acetate in methanol furnished pyrimidinone **16a.** While the yield was fairly low, it still served as proof that the intermediate biaryl **15e** was formed in the previous step. However, the following Pd-catalyzed Suzuki-Miyaura coupling with *N*-Boc-protected pyrrol-2-yl boronic acid failed to provide pyrimidinone **17**. We therefore sought an alternative synthetic path by changing the order of the chemical transformations.

Initial Pd-catalyzed coupling of biaryl **15e** to give triaryl **16b** did not only prove successful, but also provided a higher yield compared to **16a**, and the subsequent annulation step also proved to be successful. This observation is aligned with other, unpublished data from our group, which has shown that Pd-catalyzed coupling reactions fail for substrates which contain pyrimidinone rings, such as in **16a**.[155] After the reaction was worked up and purified, the deprotected final molecule **RSL-0034** was obtained instead of **17**, making the final planned Boc-deprotection step unnecessary. This "accidental" deprotection may be attributed to the formation of acetic acid combined with high temperature in the reaction mixture. The signal associated with the *tert*-butyl group in compound **16b** is clearly lacking/less intense in the ^{1}H-NMR of the crude reaction mixture (**Figure 30)**. This indicates strongly that the Boc group may have been eliminated in the course of the reaction, and combined with the fact that it has been shown that carbamate protecting groups on aromatic *N*-atoms display higher lability compared to aliphatic counterparts, this observation may not be as unexpected.[174] **RSL-0034** was purified using semi-prep RP HPLC.


OTs
CN
Br
14e
Diethyl aminomalonate hydrochloride, tBuOK, EtOH, r.t.
NH2
CO2Et
N
H
Br
15e
(Not isolated)
Formamidine acetate, MeOH, Δ
8%
over 3 steps
N
NH
O
16a
N-Boc-2-pyrrolyl boronic acid, Cs2CO3, Pd(dppf)Cl2CH2Cl2
n.r.
17, R = Boc
RSL-0034, R= H (10%)
NR
10%
over 2 steps
N-Boc-2-pyrrolyl boronic acid, Cs2CO3, Pd(dppf)Cl2CH2Cl2
20%
over 3 steps
16b
NBoc
Formamidine acetate, MeOH, Δ

Scheme 8: Synthesis of RSL-0034 from tosylate **14e.**

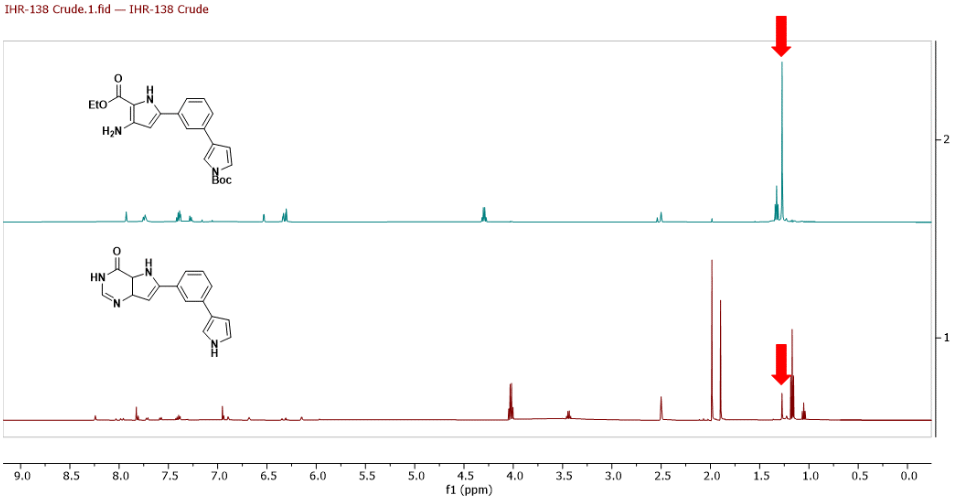
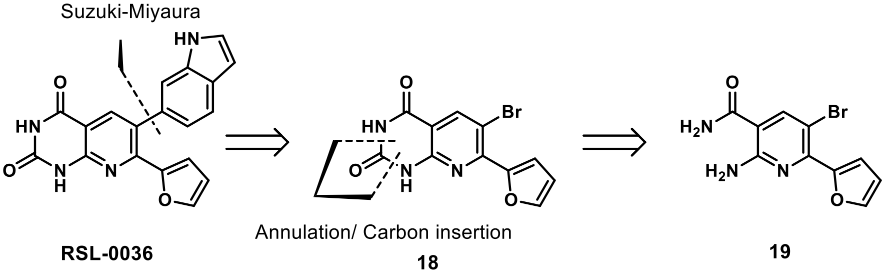


Figure 30: *Top:* ^{1}H-NMR spectrum of Boc-protected compound **16b**, with the Boc-signal emphasized. *Bottom:* The signal is clearly weaker/absent in the crude mixture after the annulation of **16b.**

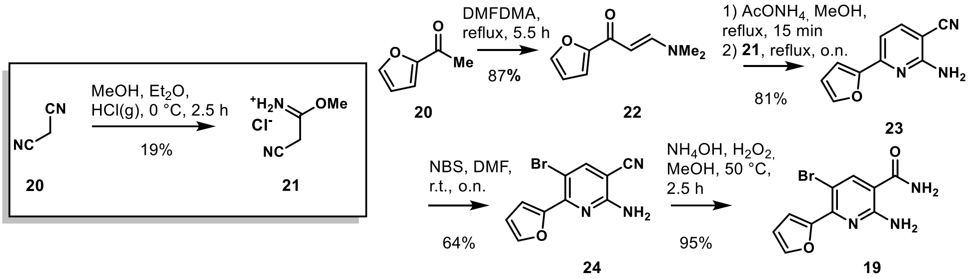
3.3.3 Synthesis of **RSL-0036**

For the synthesis of **RSL-0036**, we envisioned that we could follow the strategy developed by Lindeman *et al.* in their 7-step synthesis of a radiofluorinated ligand for PET imaging.[150] A brief outline of this strategy is shown in **Scheme 9**. The final Suzuki-Miyaura coupling from bromide **18**, which introduces the indolyl moiety in **RSL-0036**, opens up the possibility for a flexible strategy, where multiple analogues could be created in the endgame of the synthesis. Other interesting aspects of this strategy is the bypass of a Thorpe-Ziegler-like reaction, which previously proved challenging and the insertion of an extra carbonyl group to give the pyrimidinedione **18** from primary amide **19** (or alternatively, insertion of a methine group to make a pyrimidinone from primary amide **19** would also be possible). The first steps of the synthesis are summarized in **Scheme 10**.



Scheme 9: Retrosynthetic analysis of **RSL-0036**.

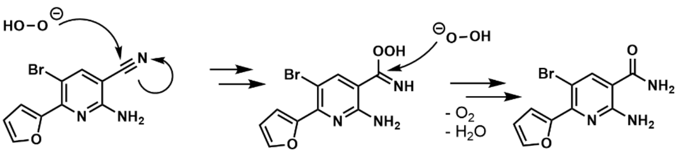
Before the construction of the main scaffold in **RSL-0036,** it was necessary to obtain the imidic acid **21**. We found that the method of hydrogen chloride gas generation played a significant role for both the yield that could be achieved and purity of the product of the reaction. The best yields were obtained when HCl gas generated by dropwise addition of concentrated sulfuric acid to dry sodium chloride was allowed to pass through a trap of concentrated sulfuric acid to dry the gas, before slowly bubbled into a solution of malononitrile **20** and methanol in dry diethyl ether.


MeOH, Et$_2$O, HCl(g), 0 °C, 2.5 h
19%
20
21
DMFDMA, reflux, 5.5 h
87%
20
22
1) AcONH$_4$, MeOH, reflux, 15 min
2) 21, reflux, o.n.
81%
23
NBS, DMF, r.t., o.n.
64%
24
NH$_4$OH, H$_2$O$_2$, MeOH, 50 °C, 2.5 h
95%
19

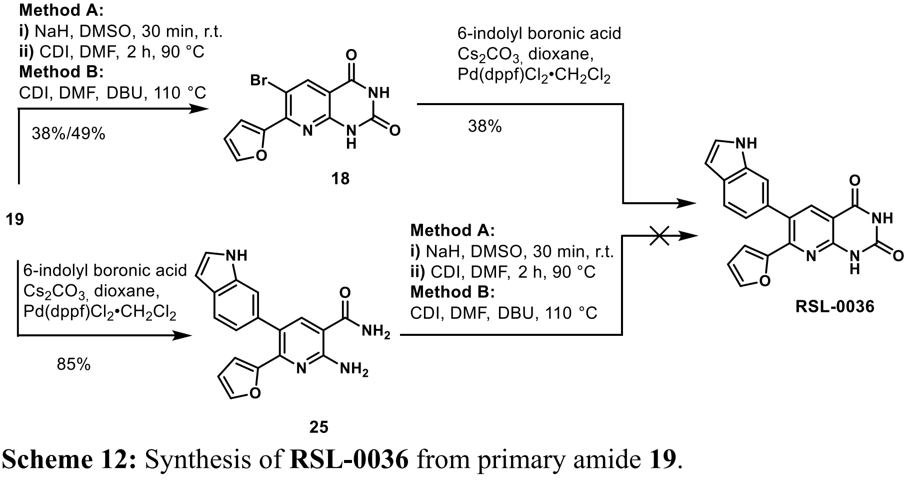
Scheme 10: Synthesis of building blocks for **RSL-0036.**

The pyridine **22** was constructed in two steps from acetyl furane **20** in good yields. The originally reported conditions for the halogenation of pyridine **23** to give bromopyridine **24** gave in our hands only a di-brominated species. By lowering the amount of NBS from two to one equivalent, bromopyridine **24** was obtained in satisfying yield. ^{1}H-NMR of the crude reaction mixture showed that the di-brominated side-product had been formed in a 4:13 ratio to the desired product. These two species were separated through means of column chromatography, but we found a small amount of the dibrominated species co-eluted with **24** (<10%)**.** The subsequent hydrolysis of nitrile **24** to primary amide **19** under the reported conditions (hydrogen peroxide, NaOH, in ethanol/H$_2$O, 50 °C, 5 h) failed to give the desired product, but rather provided the corresponding carboxylic acid. The original procedure reports the use of sodium hydroxide and hydrogen peroxide, but as a suggested mechanism of this reaction proposes one or multiple nucleophilic attacks from a deprotonated peroxyl species, (

Scheme 11) it is not unlikely that a hydroxyl moiety could take its place in this reaction.[175] Another procedure by Clayton *et al.*[176] have reported that a corresponding transformation on a similar structure had been successful using ammonium hydroxide in place of sodium hydroxide, and after submitting **24** to these conditions, amide **19** could be isolated in excellent yield (95%) with minimal purification beyond standard work-up conditions. Interestingly, Lindemann *et al.* reported having obtained **19** in only 45% yield, while also having to purify the compound using flash chromatography. Next, two different conditions involving use of carbonyldiimidazole were tested to form the pyrimidinedione ring (**Scheme 12**) of **18**, using a strong non-nucleophilic inorganic base (NaH) or a strong non-nucleophilic organic base (DBU). Both conditions proved successful in giving **18**, with a slightly better yield using DBU.


HO–O⊖
Br
N
NH2
OOH
NH
⊖O–OH
- O2
- H2O
O
NH2

Scheme 11: Proposed mechanism of the base-catalyzed conversion of ntiriles to amides by hydrogen peroxides.[175]


Method A:
i) NaH, DMSO, 30 min, r.t.
ii) CDI, DMF, 2 h, 90 °C
Method B:
CDI, DMF, DBU, 110 °C
38%/49%
19
18
6-indolyl boronic acid
Cs2CO3, dioxane,
Pd(dppf)Cl2•CH2Cl2
38%
85%
25
RSL-0036

Scheme 12: Synthesis of **RSL-0036** from primary amide **19**.

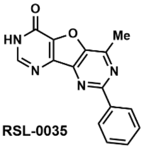
Standard Suzuki-Miyaura conditions failed to provide the target molecule **RSL-0036** from **18**, however, the same conditions proved quite successful in furnishing **25** from

19. The earlier mentioned conditions using carbonyl diimidazole for ring closure unfortunately failed to provide the target molecule in this instance as well. Stumped at how to proceed to obtain the desired final product, turning back to the original publication's conditions described that the Suzuki-Miyaura coupling was performed with 10 equivalents of base (as opposed to the standard 2 equivalents required for the metathesis and transmetalation steps).[177] This detail had initially been overlooked as it was not commented on in the original publication. Another attempt at this step with 10 equivalents of Cs_2CO_3 provided the final compound **RSL-0036** from **18** in an acceptable yield. These results echo the struggles of obtaining **RSL-0034** from intermediate **16a** (**Scheme 8**), which also failed under standard Suzuki-Miyaura conditions. It is not unreasonable to speculate that the pyrimidinone moiety complicates the matter of performing this coupling reaction successfully, either by complexation to Pd or claiming base equivalents through deprotonation. After flash chromatography on a C18 column and HPLC RP purification, the final compound **RSL-0036** was obtained in sufficient purity for testing.

3.4 Biological Data

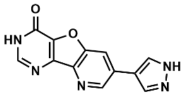
The affinity of the compounds to the FMN riboswitch was determined with a surface plasmon resonance (SPR)-based assay. This assay is a label-free biophysical experiment where a macromolecule analyte (FMN riboswitch) is immobilized on a surface (sensor chip) and the angle of reflected light upon the sensor chip is measured. The refractive property of the sensor chip is affected by the resonance of its electrons (surface plasmons), which are sensitive to small changes in the immediate environment. As such, binding of a small molecule analyte (such as FMN or an analogue) to the immobilized FMN riboswitch changes the reflection angle, and a signal is produced. This signal can be used to measure the direct affinity of a compound binding to the FMN riboswitch by calculating the dissociation constant (K_D), which is the ratio between the association rate (k_{on}) and dissociation rate (k_{off}). Alternatively, K_D can be estimated by obtaining steady state signals from different concentrations.[178] In our experiments, DNA was also immobilized on a reference sensor chip and the signal from the small molecule analyte's interactions with this immobilized DNA was subtracted from the signal obtained from the RNA/small molecule analyte interaction, in order to assess if the observed activity was specific to RNA/small molecule analyte interactions. The experiments were carried out by Dr. Fahimeh Khorsand at the Department of Biomedicine.

A summary of all the activity data and ligand efficiencies (LE) for all the compounds explored in this study are shown i**n Figure 31.**



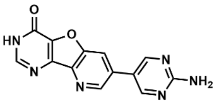
RSL-0035

Active, but poor data (Low solubility)



RSL-0029

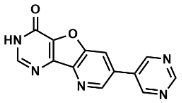
Low solubility in DMSO



RSL-0026

K_D = 2.2±0.7 μM ****

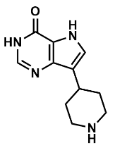
LE = 0.38 kcal · mol^{-1} · heavy atom^{-1}



RSL-0027

K_D = 5.9±0.3 μM ***

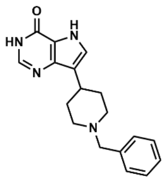
LE = 0.36 kcal · mol^{-1} · heavy atom^{-1}



RSL-0012

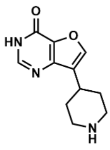
K_D = 187±60 μM ***

LE = 0.34 kcal · mol^{-1} · heavy atom^{-1}



RSL-0007

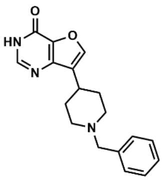
Low solubility in DMSO



RSL-0015

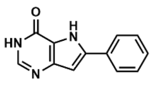
K_D = 133±18 μM ***

LE = 0.34 kcal · mol^{-1} · heavy atom^{-1}



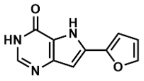
RSL-0014

Unable to fit



RSL-0031

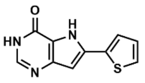
No binding



RSL-0032

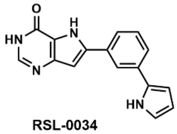
K_D = 120±30 uM ***

LE = 0.37 kcal · mol^{-1} · heavy atom^{-1}



RSL-0033

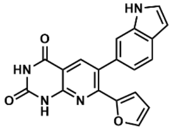
No binding



RSL-0034

K_D = 8.33±2.89 μM ***

LE = 0.34 kcal · mol^{-1} · heavy atom^{-1}



RSL-0036

Slow binding

Figure 31: Affinities from SPR-based experiments of compound explored in this thesis. Ligand efficiencies (LE) calculated from the experimental affinities Asterisks denote number of parallels performed for each compound.

The virtual screening hit **RSL-0035** displayed poor solubility in DMSO, and when assayed, data of sufficient quality could not be obtained. Considering this compound has one of the highest clogP-values of the compounds in this study (2.43, **Table 7**), this is not surprising. As such, this compound was judged to be not suitable for further work.

The closely related analogue **RSL-0029** was also troubled with solubility issues, which is somewhat surprising considering the clogP-value for this compound is not among the highest (1.45) of the explored compounds. However, both **RSL-0026** and **RSL-0027** provided a dose-dependent binding responses to the riboswitch (**Figure 32A, B**), which resulted in pleasantly high affinities (2.2±0.7 and 5.9±0.3 μM, respectively). These high activities show that while **RSL-0035** itself might be a poor candidate for further optimization, proposed analogues could still be active. When inspecting the curve data for these compounds, it is clear that both the association and dissociation rates for these two compounds compared to some of the compounds shown later are slow, which might imply the compounds have to overcome some steric hindrance in order to gain access to the binding site. It is also interesting that these two compounds displayed relatively similar affinities and LE (**RSL-0026**: 0.38 $\text{kcal}^{-1}\text{mol}^{-1}\text{heavy atom}^{-1}$, **RSL-0027**: 0.36 $\text{kcal}^{-1}\text{mol}^{-1}\text{heavy atom}^{-1}$) as the model binding modes suggested two different binding modes. As such, it is possible no contribution is made from the amine group in **RSL-0026**. It could be interesting to subject an analogue composed of the three-ringed scaffold with no substituents on ring C to the SPR assay, to see if this compound would have a similar affinity as **RSL-0026** and **RSL-0027**, and also to see if this compound would have a different kinetic profile due to lower steric hindrance.

RSL-0012 and **RSL-0015** proved to have moderate affinities of 187±60 µM and 133±18 µM, respectively (**Figure 32C, D**). Inspecting the affinity curves of these data, it is also clear that the former compound displayed a less ideal response than the latter.

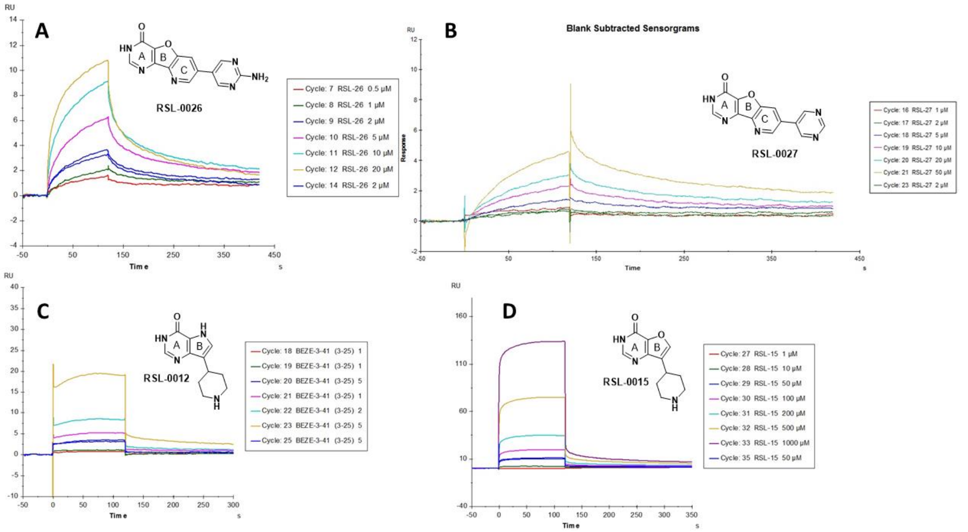


Figure 32: SPR sensorgrams for **A) RSL-0029**, **B) RSL-0027**, **C) RSL-0012** and **D) RSL-0015.**

Compared to **RSL-0026** and **RSL-0027**, it is possible the weaker affinity of these two compounds is due to the loss of ring C in the central scaffold, which results in a lower hydrophobic contact surface area between the ligand and the binding site. However, despite their weaker affinities, LE of these compounds (**RSL-0012:** 0.33 kcal^{-1}mol^{-1}heavy atom^{-1}, **RSL-0015:** 0.34 kcal^{-1}mol^{-1}heavy atom^{-1}) are still within high values. Comparing the LE between these two compounds makes it clear that even if a hydrogen bond is formed between the endocyclic nitrogen in ring B of the scaffold in **RSL-0012** and with O2 ribose backbone of A49 (as proposed in **Figure 24A**), it does not contribute significantly to affinity. **RSL-0007** and **RSL-0014** were also troubled with solubility and data fitting issues. This was surprising, as one would expect that the tertiary amine would allow for a protonation site which would ensure solubilization and as the drug-likeness indices of these compounds were among the highest (**Table 7**). It is possible the benzyl moiety which is found in these compounds is either hindering solubilization of the compound, or also that it hinders the structure from accessing the binding site due to steric hindrance. To rule out the first case, these compounds should be resubmitted for the SPR assay in the form of HCl salts to see if solubilization could be improved.

The four compounds **RSL-0031**, **RSL-0032**, **RSL-0033** and **RSL-0034** were also assayed for activity. A negative response was obtained from **RSL-0031** and **RSL-0033** (**Figure 33A, C**), which implied that these compounds have higher affinity towards the immobilized DNA used as reference for the binding curve data instead of RNA. As such, the activity of these compounds was deemed to be negative. **RSL-0032** and **RSL-0034** bind in a dose-dependent manner to the riboswitch (**Figure 33B, Figure 33D**, respectively), and the data shows faster association and dissociation rates than for **RSL-0026** and **RSL-0027**. However, curve fitting to determine a K_D value resulted in a poor fit. As such, steady state analysis of the activity was performed to determine the affinities of these two compounds (**Figure 34**). A K_D value of 120±30 µM and 8.3±2.9 µM were assigned to each of these compounds.

The following conclusions for the bicyclic scaffold can be drawn from the preceding results: a thiophenyl substituent as ring C in **RSL-0033** does not seem to be ideal for binding to the riboswitch but might instead promotes binding to DNA. A phenyl ring in the same position, such as in **RSL-0031**, by itself does not provide sufficient π/π-stacking to give suitable affinity to the riboswitch. A furanyl substituent as ring C, such as in **RSL-0032**, seems to create a favorable interaction, and is therefore able to allow specific binding to the pocket. An aromatic substituent in the *meta*-position of ring B in **RSL-0034** is favorable and allows binding.

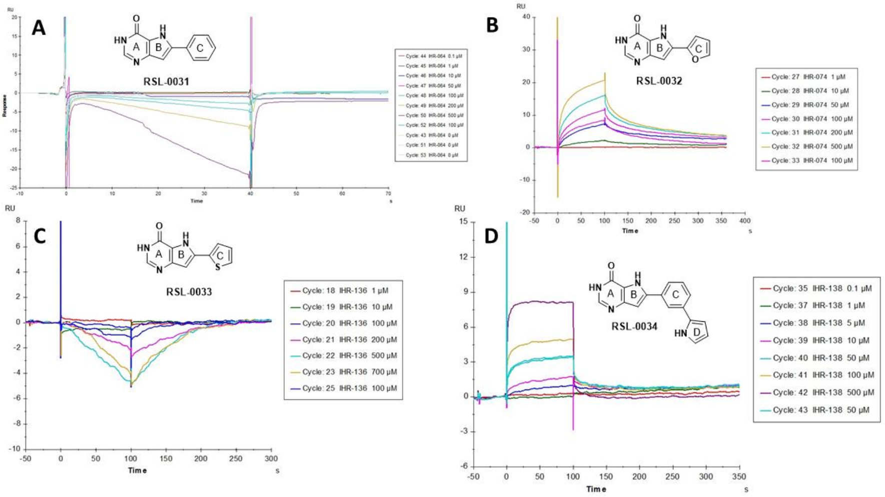


Figure 33: Sensorgrams for **A) RSL-0031, B) RSL-0032, C) RSL-003** and **D) RSL-0034.**

The observed lack of specific binding in **RSL-0031** compared to **RSL-0034** is interesting. It is likely that there is no hydrophobic contribution from the phenyl ring C in **RSL-0034**, as **RSL-0031** shows no activity and the modelling in **Figure 26D** indicates no π/π-interactions with the nucleobase A49, and that this ring merely serves as a bridge to allow the gain from a hydrophobic edge/face-interaction from ring D with G62, which is sufficient to enable the structure to gain specific binding with the FMN riboswitch pocket. It is also possible **RSL-0031** binds to RNA but shows stronger affinity to DNA. It also seems that the *para*—substitution of ring D and B on ring C also prevents the compound from binding to DNA (as seen in **RSL-0031**) and leads to specific binding of **RSL-0034** to the riboswitch.

One also has to consider that the affinity of **RSL-0034** was found to be two orders of magnitude greater than **RSL-0032**. Possible explanations for this are: either the π/π-interaction obtained between G62 and ring D in **RSL-0034** is more beneficial than the π/π-interaction obtained between A49 and ring C in **RSL-0032**, or the modelling of the binding mode for RSL-0034 is incorrect, and contacts between both ring C and D of **RSL-0034** with the binding site are formed. Considering that the ligand efficiencies of these two compounds are quite comparable (**RSL-0032**: 0.37 kcal^{-1}mol^{-1}heavy atom^{-1}, **RSL-0034**: 0.34 kcal^{-1}mol^{-1}heavy atom^{-1}), the former explanation seems likely. To answer these questions, it would be interesting to obtain a crystal structure of the FMN riboswitch/**RSL-0034** complex to elucidate the binding mode of this compound.

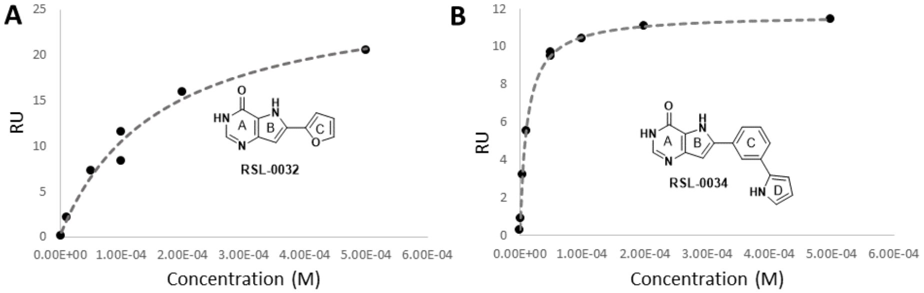


Figure 34: Steady state plots for **A) RSL-0032** and **B) RSL-0034.**

RSL-0036 was also evaluated in an SPR assay. While providing a positive, concentration-dependent response, as can be seen in **Figure 35**, the assay failed to reach a steady state within the 100 second association time. Furthermore, dissociation failed to return to a steady baseline within the >300 second timeframe. From this it seems that this compound indeed shows affinity towards the FMN riboswitch, and the long dissociation time gives hope for a high affinity. However, a proper K_D value was not obtained from this experiment.

It is possible the long association times are related to steric hindrance from ring C and D disallowing the structure to properly fit in the binding site or that a conformational change of the riboswitch is necessary to allow binding. An attempt at purifying compound **19a** was made to assess if the lack of the indolyl ring C would allow for a faster association and dissociation time, but at the time the compound had decomposed fully, and proper purification was not feasible.

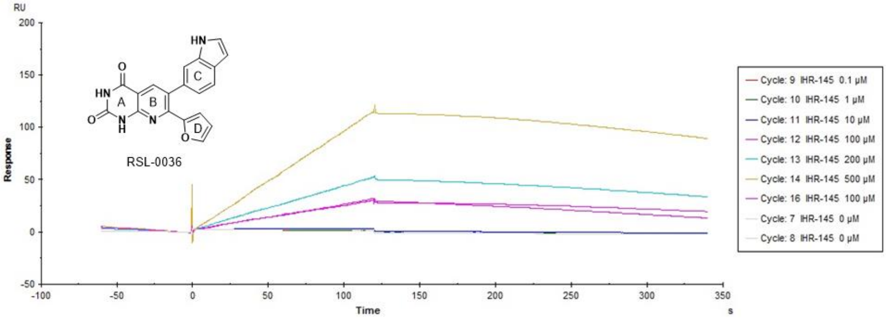


Figure 35: Sensorgram from SPR assay of **RSL-0036**.

4. Conclusions and further work

The work presented herein has focused on RNA binding sites (with emphasis on riboswitches) as potential drug targets for new classes of antibiotics.

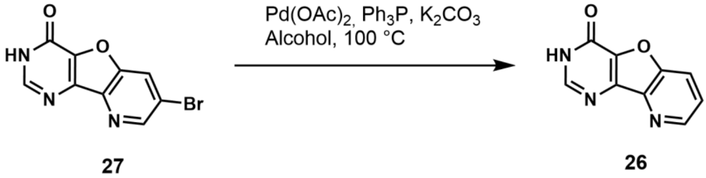
In the first part of this thesis, we focused on whether it was possible to identify RNA pockets which could be addressed with conventional drug-like ligands. This resulted in the creation of DrugPred_RNA, a structure-based druggability predictor. As RNA is an underexplored drug target, the predictor was trained on a set of protein binding sites with descriptors that could be calculated for both types of pockets. We have also shown that this predictor provides comparable quality to the previous iteration of DrugPred, which was trained on a slightly different set of descriptors (**Table 3**). Two sets of crystal structures were retrieved from the Protein Data Bank; one consisting of all available RNA-exclusive structures which contained ligands, and the other consisting of RNA-containing ribosomal structures with ligands. The predictions of anecdotally selected examples were in agreement of with manual druggability assignments of the binding sites **(Figure 16**). In addition, we have found a strong correlation between positive druggability predictions of binding sites and the occurrence of highly drug-like ligands (**Table 5**, **Table 6**). We have also shown the predictions of DrugPred_RNA to be consistent and robust towards small changes in binding site composition and conformations **(Figure 14)**, but also shown that binding sites with major different conformational states or sequencial composition could obtain different druggability predictions (Figure 18).

Over 1000 binding sites were surveyed and only 22 of them bound to a ligand described as highly drug-like and of these and 18 had reported affinity data. Near 400 sites were identified to be druggable by DrugPred_RNA (**Table 4**), which shows that there is ample untapped potential for RNA as a drug target. Among these, many riboswitches were represented, which underlines the opportunity of these as potential antibiotic targets. Many ribosomal binding sites were also identified as druggable, which can help direct efforts towards the discovery of new antibiotics. The work related to this part of the thesis has been published in *Journal of Chemical Information and Modeling.*[179]

The FMN riboswitch was among the riboswitch classes which were identified as druggable by DrugPred_RNA. On the basis of known interactions between the FMN riboswitch binding site and its native ligand FMN, and the identified highly potent binder ribocil A, we have explored the binding site and proposed new compounds (**Figure 21**) which could bind to this riboswitch. We designed and proposed 4 series of compounds which we identified as synthetically feasible and containing fragment-like properties (**Table 7**). These compounds had predicted binding modes which retained the identified key interactions between the binding site and FMN/ribocil, but at the same time explored some of the interactions inspired by ribocil or interactions not previously described (**Figure 22, Figure 24, Figure 26, Figure 27**). Some of these compounds were designed as analogues to a previously identified virtual screening hit, **RSL-0035**, while others were designed on the basis of a new central scaffold which could probe interactions in new vectors, while the final compound, **RSL-0036**, was designed as an off-shoot of recently published work, with a different central scaffold than the previously described compounds.

The syntheses of these compounds were successful, of some are described herein. Additionally, the screening hit **RSL-0035** was synthesized. The synthetic routes can serve as a basis for future work toward synthesis of analogues. Through this work, we have also identified some steps, in particular the Thorpe-Ziegler reaction, to be troublesome and low yielding **(Table 8, Table 9**). We have also identified some unusual ^{13}C-NMR behavior (**Figure 29, Table 10**) of substituted pyrrole compounds which could lay the foundation for another NMR-based study.

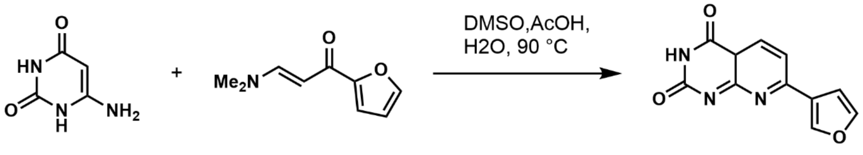
The SPR assay of these compounds showed that not only were 7 of 13 compounds active, but 3 had single digit micromolar affinity. Considering the fragment-like properties of the compounds, the affinities are remarkable and also with very promising LEs (**Figure 31**). These results are very promising for the future directions of this project. We showed that **RSL-0035** is less ideal candidate for further work, but the results from **RSL-0026** and **RSL-0027 (Figure 32A, B)** suggests that the initial proposed binding mode could still be achieved. It would be interesting to test an analogue of these compounds, compound **26**, which lacks the heterocyclic substituent. A precursor to this compound, bromide **27**, was created as a part of Zeeshan's and Burova's work,[152] and should be easily converted to **26** in a Pd-catalyzed dehalogenation reaction, as shown in **Scheme 13**.[180]



Scheme 13: Pd-catalyzed dehalogenation of **27** to give **26**

Resubmittal of **RSL-0012** and **RSL-0014** as HCl salts should be done to verify if these compounds are inactive or simply insoluble. The activities of **RSL-0012** and **RSL-0015 (Figure 32C, D)** imply that the hydrogen bonding which could potentially be formed in the former compound does not seem to affect activity, which can shape the synthetic efforts for future potential ligands for the FMN riboswitch. The findings from **RSL-0031**, **RSL-0032**, **RSL-0033** and **RSL-0034 (Figure 33, Figure 34)** show that the choice of aromatic substituents can influence activity and potential analogues could also be explored.

For **RSL-0036**, an analogue lacking the indolyl moiety (compound **28**) should be investigated to see if it could bind faster to the target (**Figure 35**). This compound could be prepared from **18** using a similar method as described above, or alternatively, a synthesis of this compound has been described from building block **22** and 6-aminouracil **29** by Churchill *et al.*[181], as outlined in **Scheme 14.**



29 **22** **28**

Scheme 14: Synthesis of **28** from furane **22** and 6-aminouracil **29.**

For all the active compounds mentioned above, it would be interesting to obtain an X-ray crystal structure to obtain detailed information on their binding modes which were proposed through the minimization studies. The activities should also be verified through an orthogonal method, such as isothermal calorimetry (ITC), or in the case for the compounds with low micromolar affinities, in a functional *in vitro* transcription/translation assay.

5. Methods and general experimental

5.1 DrugPred_RNA

The following text within this text related to the methods for DrugPred_RNA is taken from the related publication for DrugPred_RNA.

Scripts to download crystal structures from the PDB, process them, and calculate ligand and binding site descriptors were written using Python 3.6.8 with the Biopython (1.73) and RDKit (2019.09.1) libraries.[182,183]

NRDLD Set for Training and Validation.

Our NRDLD set with the most recent modifications was used to train and test a druggability predictor on protein targets.[71,184] In brief, this set contains 110 small molecule binding sites. The proteins in the set have a maximum sequence similarity of 60% to each other, and 68 of the binding sites were previously annotated to be druggable and 42 to be less druggable based on data mining and available literature. This set was split into a training set containing 75 pockets (47 druggable/28 less druggable) and a test set containing 35 pockets (21 druggable/14 less druggable) as done before. The binding sites and surrounding residues were carved out of the CIF files downloaded from the PDB by keeping all residues with an atom within 15 Å of the ligand to reduce the file size. The isolated parts of the structures together with co-factors and metal ions if present were saved in the PDB format and used for generating the NRDLD Set for Training and Validation. Our NRDLD set with the most recent modifications was used to train and test a druggability predictor on protein targets.[71,184]

Superligand generation

A superligand as a negative print of the binding site was obtained as done previously with minor modifications.[71] In brief, a set of approved drug molecules was docked into the pocket that contained the bound ligand in the original crystal structure using DOCK

3.6.[185] Since the aim of docking was solely to obtain information about the shape and the volume of the binding site, all receptor atoms were set to carbon atoms and assigned a partial charge of 0. Subsequently, compounds for which a docking pose was obtained and for which the ratio of van der Waals (VDW) score to number of heavy atoms was ≤ -1.3 were merged into a superligand. This cutoff was chosen to ensure that only ligands that filled the pocket were kept. To minimize the number of atoms in the final superligand, during the merging process, only atoms adhering to all of the following criteria were retained: (1) the atom had to be a nonhydrogen atom, (2) at least two atoms coming from different docked compounds had to be closer than 1.2 Å, and (3) only one of the atoms within 1.2 Å from other atoms was kept. If no docked ligands passed these filters, the ligand contained in the original complex structure was used as the superligand. This was the case in 125 instances in the RNA data set and 342 instances in the ribosomal data set.

Descriptor calculation

The binding site and buried superligand atoms were determined based on the superligand. For that purpose, using FreeSASA[186] as implemented in RDKit, the solvent accessible surface area (SASA) of each receptor and superligand atom in the superligand-bound and -unbound state was calculated using a 1.0 Å probe radius and ProtOr radii.[187] All receptor atoms for which the SASA differed between superligand-bound and -unbound state were assigned as being binding site atoms. Further, the SASA of all superligand atoms in the free state was calculated. Superligand atoms with a SASA >0 were assigned as surface atoms, and those with a SASA = 0 were assigned as buried superligand atoms. Using superligand and binding site atoms as input, descriptors describing the size, shape, and polarity of the pocket were calculated (Table S1). For shape descriptors that are not based on the surface area or the number of receptor or superligand atoms, the Descriptors3D module of RDKit was used. For calculating polarity descriptors, we considered all carbon, phosphor, and sulfur atoms in addition to nitrogen atoms of the RNA bases that are bound to the ribose to be hydrophobic and all oxygen atoms of amino acids, ribose sugars, and phosphate groups in addition to nonaromatic nitrogen atoms of amino acids to be polar. The SASA values of these atoms were calculated with FreeSASA using the same settings as described above. The side chains of histidine and tryptophane residues as well as the RNA bases are known to form hydrogen bonds in the plane of the heterocycles, while parallel to this plane, they engage in π-stacking interactions that are more hydrophobic in nature. To account for this ambivalent behavior, the SASA of endocyclic aromatic nitrogen atoms of the bases and amino acid side chains and exocyclic oxygen and nitrogen atoms of the bases was split into a hydrophobic and a polar contribution in the following way. The SASA of these atoms was calculated in both the absence (SASA_total) and the presence (SASA_pol) of two blocking carbon atoms that were placed perpendicular to the plane of the aromatic ring with a 1.70 Å distance from the atom of interest. The area SASA_pol was considered to belong to a polar atom, while the difference SASA_total − SASA_pol was considered to belong to a hydrophobic atom. Similarly, if more than half of the SASA of an atom was deemed to be hydrophobic, the atom was included in the hydrophobic binding site atom count.

Training the Predictive Model Using Decision Trees.

Machine learning was carried out using the XGBoost[134] package in R,[188] a scalable machine learning system for tree boosting. In brief, the method is based on initially creating multiple decision trees that are evolved over time into a model with increased predictive power. As a learning objective, logistic regression for binary classification with output probability was used. Thus, all binding sites obtained a score between 0.0 and 1.0, whereas pockets with a score ≥ 0.5 were labeled druggable and pockets with a score < 0.5 were labeled as less druggable. Divergent from the default settings, the following parameters were used for training the model:

- Max_depth = 3 (maximum depth of trees)

- Scale_pos_weight = 0.63 (adjusts for the skewness between druggable and less druggable binding sites in the training set)
- Early_stopping_rounds = 20 (Validation metric needs to improve at least once in every 20 rounds to continue training.)

The influence of the descriptors on the model was evaluated with the help of Shapley Additive Explanation (SHAP) values as implemented in the SHAPforxgboost package.[135,189,190] The same package was also used to make **Figure 12** and **Figure S3** in the supporting information of the related publication.[179] SHAP values describe the importance of each descriptor for the model output taking into account the interactions with other descriptors. Each descriptor for each data point (here, a particular binding site) is assigned a positive or negative SHAP value describing the contribution of the descriptor to the model output (here, druggable or less druggable) for that data point. The mean SHAP value formed by all SHAP values for a descriptor for the entire data set indicates the importance of the descriptor for the model (the larger the absolute mean SHAP value, the more important the descriptor). For DrugPred_RNA, positive SHAP values imply a high druggability probability, while negative SHAP values imply a low druggability probability. Further, by plotting the individual SHAP values for a descriptor against the descriptor values, it becomes evident which descriptor values contribute positively or negatively to the model. The sum of the SHAP values of all descriptors for a single data point indicates the direction of the prediction for that data point. Descriptors included in the final model were chosen by iteratively removing the least impactful descriptors until the predictive performance of the model was negatively affected. To further assess the robustness of the final model (called DrugPred_RNA), leave-one-out cross-validation was carried out, yielding a training and testing error of 0.00342 and 0.127, respectively.

In addition, accuracy precision and recall values of the models were calculated using eqs 1−3 with true positives (tp) and true negatives (tn) being the number of correctly classified binding sites and false positives (fp) and false negatives (tn) being the number of wrongly classified binding sites.

Equation 1:

$$Accuracy = \frac{tp + tn}{tp + fp + tn + fn}$$

Equation 2:

$$Precision = \frac{tp}{tp + fp}$$

Equation 3:

$$Recall = \frac{tp}{tp + fn}$$

Assembly of the dataset with RNA binding sites.

We selected RNA structures for druggability assessment by querying the PDB for structures containing only RNA and ligands (accessed November 2019). In addition, the PDB was searched for entries containing ligands and the keyword ″riboswitch″ to include structures that were excluded in the first query due to the presence of proteins. In total, this yielded 1084 structures. Subsequently, all structures that contained only ligands that were detergents, buffer salts, or crystallization components were filtered out, reducing the data set to 427 unique entries, see the supplementary material for the publication for three-letter codes of rejected ligands).[179]

If a crystal structure contained several instances of the same ligand, only the first instance was retained. In addition, all metal ions and water molecules were deleted (for a list of metal abbreviations, see the supplementary material). This resulted in 465 distinct binding sites spanning 224 unique ligands. A second variant of this set was also prepared. In this variant, only pockets with metal ions that were not more than 5 Å away from a ligand atom were retained. If a binding site contained several metal ions, several copies of the binding sites, each of them containing one of the metal ions, were prepared. This variant contained 343 entries. In the following, the first variant is called the metal-free and the second variant the metal-containing set. Further, a data set containing ligand binding sites in ribosome crystal structures was compiled by querying the PDB for structures that contained ″ribosome″ as a keyword. These structures were treated as described above. In addition, the ligands were visually inspected to remove buffer components that had slipped the filter rules. This resulted in 613 binding sites in the metal-free ribosome set and 546 in the metal-containing set spanning 217 unique ligands. The binding site regions were carved out of the original CIF files by keeping all RNA residues with at least one atom within 15 Å of the ligand and potentially metal ions as described for the NRDLD set and subjected to descriptor calculations.

Determination of overall sequence similarity and binding site similarity

To investigate the robustness of DrugPred_RNA toward changes in the binding site composition or conformation, binding sites were grouped into families based on overall sequence similarity and binding site similarity. For the grouping based on overall sequence similarity, the chains were aligned pairwise using BioPython’s pairwise2 global alignment function and the sequence similarity was calculated. If this value was >98%, the structures were assigned to the same family. For clustering based on binding site similarity, the binding site sequence of each pocket was generated by including all residues that contained at least one binding site atom in ascending order, while for modified nucleic residues, the name of the corresponding unmodified residue was used (see the supplementary information for the related publication for a list of residue IDs for modified residues[179]). Subsequently, all binding site sequences were aligned as described above. If the sequence similarity was >85%, the pockets were assigned to the same family.

Consensus Scoring

As done previously, the consensus of the druggability predictions within each family of sequences (C) was calculated using the following formula:

$$C = \frac{|n_d - n_{ld}|}{N} \times 100\%$$

where n_d is the number of druggable binding sites within the family, n_{ld} is the number of less druggable binding sites, and *N* is the total number of family members.23 Thus, 100% consensus would be obtained if all pockets in one family were predicted to belong to the same class (druggable or less druggable) and 0% if one half of the pockets was predicted to belong to one class and the other half to the other class.

Calculation of Ligand Properties. Physicochemical properties of the ligands in the RNA sets were calculated using RDKit. Further, the drug-likeness of ligands was estimated using the quantitative estimate of drug-likeness (QED) score as implemented in RDKit using average descriptor weights.[151] This score weighs multiple molecular features (e.g., molecular weight, number of hydrogen bond donors or acceptors, polar surface area, and presence of unwanted functionalities) into one single unitless score, which ranges from 0 (undesirable) to 1 (desirable). Although this metric does not provide a clear cutoff to distinguish “desirable” from “undesirable” compounds, the authors denoted a mean score of 0.67 for attractive compounds, 0.49 for less attractive compounds, and 0.34 for too complex and unattractive compounds. Accordingly, we classified compounds with a QED score ≥ 0.67 as drug-like, those with a QED score ≤ 0.49 as less drug-like, and those with a score in between as moderate drug-like.

5.2 Molecular Minimization and Ligand Design

Molecular modeling was performed using the Maestro package (version 12.2.012, MMshare Version 4.8.012, Release 2019-4, Platform Windowsx64).

As receptor, the structures of the FMN riboswitch in complex with FMN (PDB id: 2yie). The structure was prepared using standard parameters (bond orders assigned, hydrogen atoms added, protonation states generated at pH 7.0±2.0 using Epik). A global minimization was performed on the structure using the MacroModel minimization tool with the OPLS3e force field with standard parameters (solvent: water, charges from forcefield, PRCG method with 2500 iterations in gradient conversion mode, convergence threshold of 0.05).[191]

The natural ligand was removed, and ligand candidates were built in place using the 3D builder tool or initial binding modes were generated using docking (see below). Compounds with basic functional groups were built as the protomer which corresponded to physiological pH. All atoms within the candidate molecule and nearby residues within 6 Å were subsequently minimized using the parameters described above, and the predicted pose was visually inspected. If the compounds failed to provide the key interactions observed for ribocil and postulated for **RSL-0035** (See Introduction), it was discarded.

5.3 Docking

Compounds **RSL-0026**, **RSL-0027** and **RSL-0027** were evaluated through docking. All other compounds, except **RSL-0035**, were initially conceived through *in silico* design and subsequent minimization.

Structures were constructed *in silico* using Maestro and exported as SMILES. Possible protomers, tautomeric states and isomers were generated using in-house scripts based on the OEToolKit (OpenEye, Santa Fe NM). Three-dimensional coordinates and low energy conformations of the small molecules were generated using OMEGA2.[192] (OpenEye, Santa Fe, NM). Receptor preparation and validation of the system were initially carried out using DOCK 3.6[193], with the following sampling parameters:

- Receptor and ligand bin size were set to 0.5 Å
- Receptor and ligand overlap bin size were set to 0.4 Å
- The distance tolerance of matching pairs of receptor and ligand spheres were set to 1.2 Å
- The number of energy minimizations for the docked molecules was 20

The phosphate groups were treated as neutral, to balance the overall charge as done before.[194] The docked molecules were ranked according to a scoring function, which contains terms related to electrostatic and vdW interaction energies as well as correction for desolvation energies of the ligands:

$$E = E_{elec} + E_{vdW} - \Delta G_{solv}$$

where E_{elec} represents the electrostatic interaction energy term, E_{vdW} the van der Waals contribution and ΔG_{solv} the desolvation energy for the ligand when moving from the bulk solvent to the binding site. The docking was visually inspected and only poses matching expected binding modes were selected for further evaluation, using molecular minimization (see above).

5.4 General Synthetic Experimental Methods

Chemicals and solvents were purchased from Sigma-Aldrich and used as delivered unless otherwise stated. All moisture sensitive reactions were carried out under argon atmosphere in oven-dried (130 °C) equipment that has been cooled down under vacuum.

Anhydrous THF was obtained from an anhydrous solvent delivery system (SPS-800 system from M. Braun GmbH, Garching, Germany).

Thin layer chromatography (TLC) analyses were performed on aluminium sheets coated with Merck TLC silica gel 60 F254 and visualization was achieved by using ultraviolet light (254 nm).

Flash column chromatography was performed either manually using silica gel from Merck (Silica gel 60, 0.040–0.063 mm) or was performed on a PuriFlash XS 420 system (Interchim, Montlucon Cedex, France) using pre-packed columns as specified for each compound.

Preparative high-performance liquid chromatography (HPLC) was performed on a Thermo 321 multisolvent pump with an UltiMate 3000 variable wavelength detector using an XBridge Prep C18 5 µm OBD (19 x 250 mm, 5 µm) column with mixtures of acetonitrile and water (both containing 0.1% TFA) as eluent.

Analytical HPLC was performed on a 1290 Infinity II Flexible pump with a 1260 Infinity II DAD WR detector using a ZORBAX RRHD Eclipse plus 300-SB C18 (50 x 2.1 mm, 300 Å, 1.8 µm) column with mixtures of acetonitrile and water (both containing 0.1% TFA) as eluent.

The NMR experiments were recorded on a Bruker BioSpin AV500 or a Bruker BioSpin Ascend spectrometer operating at 850 MHz for ^{1}H equipped with an inverse-detected triple resonance (TCI) cryoprobe. ^{1}H and ^{13}C chemical shifts (δ) are reported in ppm with reference to the solvent residual peak ($CDCl_3$: δH 7.26 and δC 77.16; $(CD_3)_2SO$: δH 2.50 and δC 39.98). All coupling constants are given in Hz.

5.5 Activity assay

The SPR experiments were carried out by Dr. Fahimeh Khorsand, Department of Biomedicine, University of Bergen. The data were obtained via SPR using a Biacore T200 (GE Healthcare Life Sciences). Briefly, biotinylated FMN-riboswitch RNA was captured on a NAHC 1500M chip; XanTec Bioanalytics GmbH to an immobilization level of around 2000 RU at a flow rate of 10 µl/min. Biotinylated synthetic DNA was used as the reference. The running buffer for the immobilization of RNA and DNA was PBSP (GE Healthcare Life Sciences) including 2mM $MgCl_2$. The binding assay of the tested molecules was carried out using the same buffer containing 5% DMSO at the flow rate of 30µl/min at 25 °C. To measure the association rate, different concentrations of studied fragments were injected for 80 to 120 seconds over the flow cells and dissociation was initiated with the injection of running buffer which continued for 120 to 300 seconds. The Biacore T200 Evaluation Software 3.0 was applied for the evaluation of steady state affinity and all the kinetic data using a global fit and assuming a 1:1 Langmuir binding model.

5.6 Other

5.6.1 Molecular descriptors of *in silico* designed compounds in Table 7

ClogP-values based on the Crippen approach[195], QED-values[151] and number of hydrogen bond acceptors/donors were calculated using RDKit (2019.09.1).

5.6.2 HRMS analyses

HRMS-analyses were performed by Dr. Bjarte Holmelid on a JMS-T100LC AccuTOF from JEOL, USA, Inc. (Peabody, MA, USA). Orthogonal accelerated time of flight single stage reflectron mass analyzer equipped with a dual micro channel plate detector.

The AccuTOFTM mass spectrometer operated in the positive/negative mode at a resolving power of approximately 6000 FWHM. The atmospheric pressure interface zone had three different operation conditions (method 1, method 2 and method 3, see below). The TOF mass selection window was set to detect m/z values up to 2000 and the mass spectral acquisition settings applied were as follows; spectral recording interval = 0.5 s, wait time = 0.03 ns and data sampling interval = 0.5 ns. The samples were analysed as solutions in methanol or acetonitrile (~ 50 mg/mL) and introduced to the ESI spray chamber by either methanol, acetonitrile or a mixture of acetonitrile and water (50:50) used as spray reagent. Internal mass calibration was performed using a 10 ppm solution of PEG600 (polyethylene glycol average mass 600 u) in methanol (ESI+) or 10 ppm solution of NaTFA in methanol (ESI-) recorded in the same acquisition as the chemical sample. The total ion chromatogram (TIC) was acquired for approximately 2 min and the data were processed by creating extracted ion chromatograms with an m/z interval of ±0.1 u. The mass spectrums were calibrated using the mass spectrum of PEG600 or NaTFA acquired during the same set of experiments.

Three methods were used for HRMS data acquisition. The relevant compounds for each method are listed in paranthesis:

Method 1 (Used for compound **15b, 15c, RSL-0036, RSL-0034, 16a, 16b** and **19**):

ESI positive mode. Needle voltage: 2200 V, Orifice 1 = 25-20 V, Orifice 2 = 8-6 V and ring lens = 12-10 V. The temperature of orifice 1 was kept at 100 °C. The voltage of the ion guide (peak to peak voltage) was varied between 1000 and 2500 V to apply transmission of ions of different m/z ratios.

Method 2 (Used for compound **5, 6, RSL-0031, RSL-0033, 7, RSL-0032, 2a** and **2b**):

ESI positive mode. Needle voltage: 2600 V, Orifice 1 = 12 V, Orifice 2 = 6 V and ring lens = 27 V. The temperature of orifice 1 was kept at 100 °C. The voltage of the ion guide (peak to peak voltage) was varied between 1000 and 2500 V to apply transmission of ions of different m/z ratios.

Method 3 (Used for **RSL-0035**):

ESI negative mode. Needle voltage: -3200 V, Orifice 1 = -12 V, Orifice 2 = -6 V and ring lens = -27 V. The temperature of orifice 1 was kept at 100 °C. The voltage of the ion guide (peak to peak voltage) was varied between 1000 and 2500 V to apply transmission of ions of different m/z ratios.

5.6.3 NMR predictions in Table 10

The following parameters were used for incremental ^{13}C shift prediction calculations if available: Frequency was set to 500 MHz and solvent to DMSO. For ChemDraw (© PerkinElmer Informatics, Inc.), version 19.1.1.21 was used. NMRDb.org and NMRShiftDB was calculated in browser.[171,172] For NMRShiftDB, 6 spheres were used for the calculations.

Quantum mechanical-based ^{13}C shift predictions were performed by Associate Professor Nils Åge Frøystein, Department of Chemistry, University of Bergen using the software NWChem[173]. The structure of **15b** was minimized in gas phase to the lowest energetic conformation prior to shift predictions. Parameters used were according to the recommendations of Tantillo *et al.*[196,197], using the hybrid functional B3LYP/6-31+G(d,p) basis set for molecular geometric energy minimization, the B3LYP+6-31+(d,p) combination for computing isotropic shielding constants, slope: -0.9600 and intercept: 31.6874 as scaling factors for the linear regression model, and the COSMO solvation model to account for solvent (DMSO) effects.

MAE (mean absolute error) was calculated using the following formula:

$$MAE = \frac{\Sigma(|\delta_E - \delta_P|)}{N}$$

Where δ_E is the experimental ^{13}C shift value, δ_P is the predicted ^{13}C shift value and N is the number of ^{13}C signals.

5.6.4 Synthetic experiments

The experimental procedures for the synthesis of the compounds **RSL-0035, RSL-0031, RSL-0032, 7, RSL-0033, RSL-0034** and **RSL-0036** are described in the following experimental chapter.

5.6.5 Graphics software

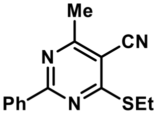
All synthetic schemes and IUPAC names of the compounds in the experimental section were drawn using ChemDraw (© PerkinElmer Informatics, Inc.), version 19.1.1.21.

Figures depicting crystal structures and the *in silico* designed compounds were made using PyMOL Molecular Graphics System v 2.3.3 © Schrodinger LLC

6. Experimental

6.1 Experiments leading up to and including **RSL-0035**

4-(Ethylthio)-6-methyl-2-phenylpyrimidine-5-carbonitrile (**6**)

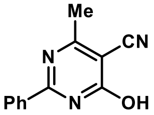


Step 1: 3-Phenylbut-2-enenitrile (5.05 g, 60.9 mmol) and benzoyl isothiocyanate (9.93 g, 60.9 mmol) were dissolved in dry acetone (80 mL) and the resulting mixture was heated to reflux for 4 h. The resulting dark orange mixture was evaporated *in vacuo* and the residue was used directly in the next step without further purification.

Step 2: To a solution of the crude product from the previous step in dry ethanol (80 mL), K_2CO_3 (8.41 g, 60.9 mmol, 2 eq.) and ethyl bromide (6.52 g, 60.3 mmol, 1 eq.) were added. After stirring under reflux for 2 h, another portion of K_2CO_3 (8.41 g, 60.9 mmol) and ethyl bromide (6.52 g, 60.3 mmol) were added, and the reaction mixture was heated to reflux for 2 h before stirring was continued overnight at room temperature. The product that was formed was isolated by precipitation and filtrated twice on a Büchner funnel and washed with a cold solution of water and ethanol (500mL, 1:1) to give the title compound as an orange solid (11.8 g, 75% over two steps).

^{1}H-NMR (500 MHz, $(CD_3)_2SO$): δ = 8.45-8.44 (m, 2H), 7.63-7.57 (m, 3H), 3.42 (q, *J* = 7.4, 2H), 2.64 (s, 3H), 1.42 (t, *J* = 7.4, 3H); ^{13}C-NMR (125 MHz, $(CD_3)_2SO$): δ = 172.5, 169.8, 162.4, 135.6, 132.4, 128.9, 128.7, 114.6, 102.7, 24.1, 23.3, 14.1; HRMS (ESI) *m/z* calcd for $C_{14}H_{14}N_3S$ $[M+H]^+$: 256.0908; found: 256.0903.

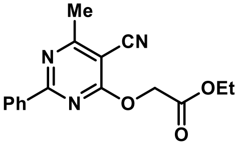
4-Hydroxy-6-methyl-2-phenylpyrimidine-5-carbonitrile (**7**)



Thioether **6** (6.63 g, 30.0 mmol) was dissolved in absolute ethanol (100 mL) and a 1.0 M solution of NaOH (50 mL) was added and the resulting mixture was stirred overnight at room temperature after which TLC analysis showed full conversion of the starting material. After the reaction was over, the reaction was subjected to a gentle nitrogen gas flow, while passing the exhaust through a bleach solution to neutralize ethyl thiol gas odor. After 15 minutes, the gas flow was removed. The solution was filtered on a Büchner funnel to give the title compound as a beige/white solid (5.25 g, 95%).

^{1}H-NMR (500 MHz, MeOD) δ = 8.09 (d, *J* = 8.0, 2H), 7.67-7.63 (m, 1H), 7.58-7.55 (m, 2H), 2.61 (s, 3H); ^{13}C NMR; Due to poor solubility, a proper spectrum was not obtained. Remaining analytical data matches previously reported data.[156]; HRMS (ESI) *m/z* calcd for $C_{12}H_{10}N_3O$ [M+H]$^+$: 212.0824, found: 212.0819

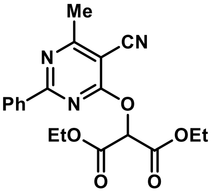
Ethyl 2-((5-cyano-6-methyl-2-phenylpyrimidin-4-yl)oxy)acetate (**2a)**



To a solution of hydroxy-pyrimidine **7** (504 mg, 2.39 mmol) and DBU (749 mg, 4.85 mmol) in dry THF (5.00 mL), ethyl bromoacetate (452 mg, 2.71 mmol) was added dropwise over 15 minutes. After stirring overnight at room temperature, the reaction mixture was partitioned between water (50 mL) and EtOAc 3 x 50 mL). The combined organic phases were dried using $MgSO_4$, and after solvent was removed *in vacuo*. The crude material was purified using a Biotage Sfär silica gel cartridge (20 µm, 25 g) eluting with EtOAc in hexane (1% to 20% over 30 min and 30 mL/min flow rate). The peak that eluted at approx. 9% was collected. Removal of the solvent under reduced pressure gave the title compound as a yellow solid (126 mg, 36%)

^{1}H-NMR (500 MHz, $CDCl_3$): δ = 8.40-8.38 (m, 2H), 7.55-7.52 (m, 1H), 7.49-7.46 (m, 2H), 5.08 (s, 2H,), 4.25 (q, *J* = 7.1, 2H), 2.75 (s, 3H), 1.27 (t, *J* = 7.1, 2H); ^{13}C-NMR: (125 MHz, $CDCl_3$) δ = 173.0, 168.7, 167.5, 164.9, 135.7, 132.5, 129.2, 128.8, 113.7, 92.8, 63.8, 61.8, 23.7, 14.3; HRMS (ESI) *m/z* calcd for $C_{16}H_{16}N_3O_3^+$ $[M+H]^+$: 298.1192; found: 298.1189

Diethyl 2-((5-cyano-6-methyl-2-phenylpyrimidin-4-yl)oxy)malonate (**2b)**

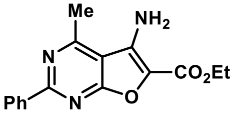


Method A: To a solution of hydroxy-pyrimidine **7** (505 mg, 2.39 mmol) and DBU (747 mg, 4.85 mmol) in dry THF (5.00 mL), diethyl bromomalonate (637 mg, 2.67 mmol) was added dropwise over 15 minutes. After stirring overnight at room temperature, the reaction mixture was partitioned between water (50 mL) and EtOAc 3 x 50 mL). The combined organic phases were dried using $MgSO_4$, and after solvent was removed *in vacuo*. The crude material was purified using a Biotage Sfär silica gel cartridge (20 µm, 25 g) eluting with EtOAc in hexane (1% to 20% over 30 minutes and 30 mL/min flow rate). The peak that eluted at approx. 18% was collected. Removal of the solvent under reduced pressure gave the title compound as an orange solid (252 mg, 28.6%)

Method B: Hydroxy-pyrimidine **7** (822 mg, 2.22 mmol) was dissolved in ethanol (2 mL), and potassium *tert*-butoxide (384 mg, 3.41 mmol) dissolved in ethanol (2 mL) was added dropwise over 15 minutes. After stirring at 10 minutes at room temperature, TLC analysis indicated full conversion and the reaction mixture was quenched with 4 M HCl (as indicated by pH paper). The reaction mixture was extracted with EtOAc (4 x 50 mL) and water (50 mL). The combined organic phases were dried over Na_2SO_4. After removal of solvent, the crude material was isolated as an orange solid and used in the next step without further purification (325 mg, 49% yield)

^{1}H-NMR (500 MHz, $CDCl_3$): δ = 8.39-8.38 (m, 2H), 7.56-7.53 (m, 1H), 7.50-7.47 (m, 2H), 5.93, (s, 1H), 4.34 (q, 4H, J = 7.1), 2.78 (s, 3H), 1.31 (t, 6H, J = 7.1); ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 173.3, 167.5, 164.7, 164.0, 135.3, 132.6, 129.2, 128.7, 113.0, 92.9, 74.4, 62.8, 23.6, 14.0; HRMS (ESI) *m/z* calcd for $C_{19}H_{20}N_3O_5{}^+$ [M+H]$^+$: 370.1403, found: 370.1400

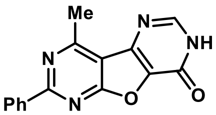
Ethyl 5-amino-4-methyl-2-phenylfuro[2,3-d]pyrimidine-6-carboxylate (**1**)



To a MW-tube was added diethyl malonate ester **2b** (53.5 mg, 0.144 mmol) which was then sealed and purged with Ar. While immersed in an ice bath, a solution of sodium ethoxide (4.4 mL, 0.1 M) was added dropwise over 10 minutes, and then immediately the temperature was increased to 80 °C for 20 min. The content of the vial was emptied into ice-cold water (30 mL) and extracted with EtOAc (3 x 30 mL). The combined organic layers were dried over $MgSO_4$, and the solvent was removed, giving around 62 mg of crude. The product was purified on a silica gel column (Hx:EtOAc, 7:3) and the title compound was isolated as a yellow solid (16.6 mg, 33% yield).

^{1}H-NMR (500 MHz, $CDCl_3$): δ = 8.54 (resolution shows only one peak for these signals even though multiplet) (m, 2H)), 7.49 (m, 3H), 5.21 (s, 2H), 4.42 (q, 2H, *J* = 6.8 Hz), 2.89 (s, 3H), 1.43 (t, 3H, *J* = 6.8); ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 165.9, 165.1, 163.5, 163.0, 161.5, 137.1, 131.3, 128.8, 128.7, 124,1 109.7, 60.9, 22.5, 14.6; HRMS (ESI): *m/z* calcd for $C_{16}H_{16}N_3O_3^+$ $[M+H]^+$: 298.1192, found: 298.1189

9-methyl-7-phenylfuro[2,3-d:4,5-d']dipyrimidin-4(3H)-one (**RSL-0035**)

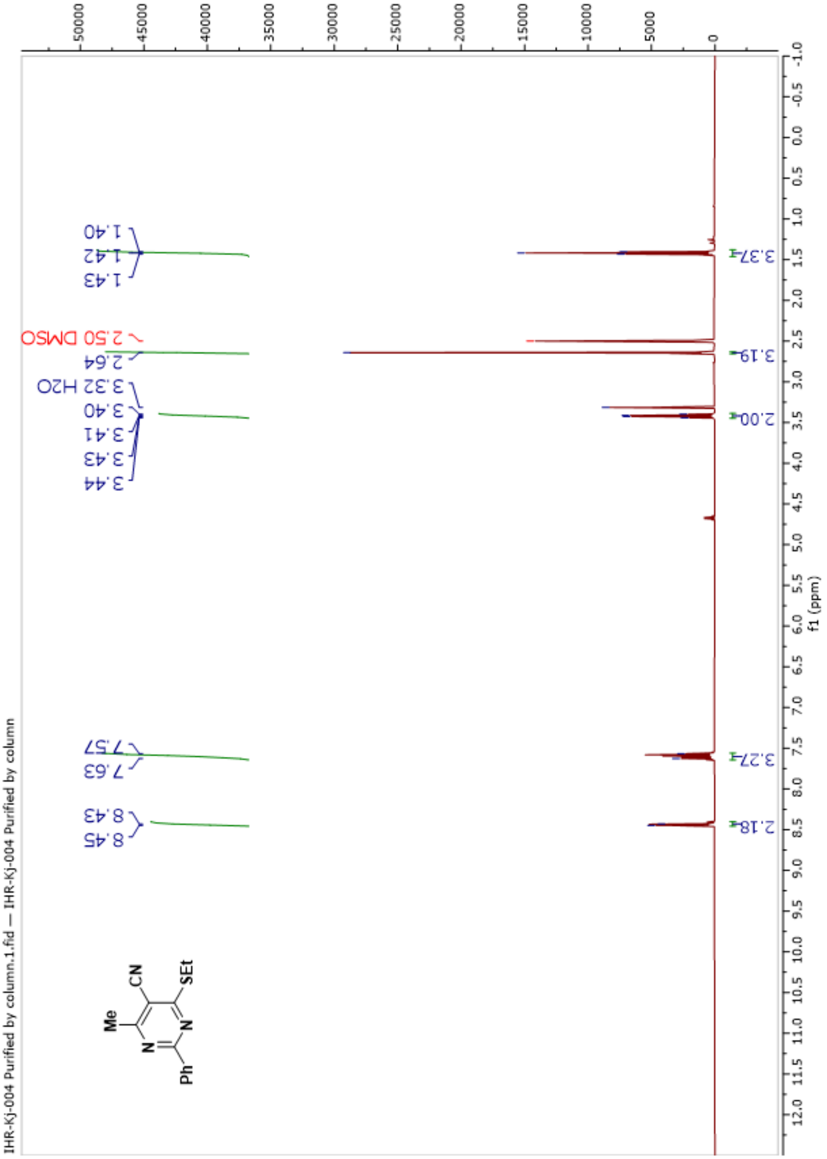


To a MW-tube was added amine **7** (0.325 g, 1.09 mmol) and formamide (20 mL). The solution was purged with an argon gas flow for 15 minutes, before formamidine acetate

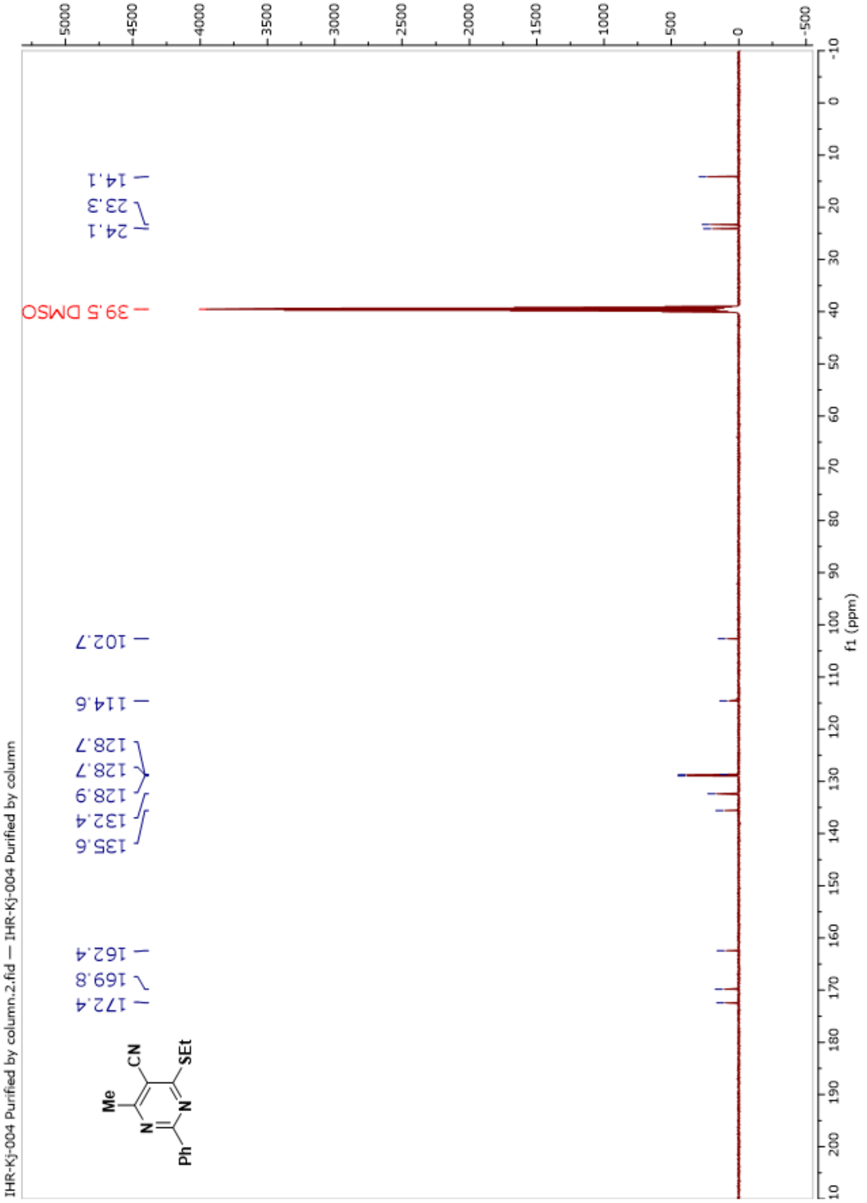
(1.33 g, 12.8 mmol) was added. The tube was heated to 150 °C and left stirring overnight for a total of 12 hours. The mixture was extracted with EtOAc (3 x 50 mL) and water (50 mL). The crude material was purified using a Biotage Sfär C18 cartridge (30 µm, 12 g) eluting with EtOAc in hexane (10% to 60% over 20 min and 30 mL/min flow rate). The peak that eluted at approx. 9% was collected. The product was further purified which was purified on a XBridge Prep C18 5 µm OBD column using a 40-80% H_2O/acetonitrile + 0.1% TFA gradient. The collected peak eluted around 70% acetonitrile (14 min). The combined fractions were concentrated under reduced pressure anfd lyophilized to give 10.1 mg of final product (0.0363 mmol, 19%).

^{1}H-NMR (500 MHz, $(CD_3)_2SO$): δ = 13.21, (s, 1H), 8.51-8.50 (m, 2H), 8.37 (s, 1H), 7.60-7.58 (m, 3H), 3.02 (s, 3H); 13C-NMR (125 MHz, $(CD_3)_2SO$)): δ = 167.6, 165.3, 162.6, 151.9, 148.1, 141.9, 137.8, 136.3, 131.6, 128.9, 128.4, 110.3, 21.9; HRMS (ESI) $[M+H]^-$: *m/z* calcd for $C_{15}H_9N_4O_2^+$: 277.0726, found: 277.0723

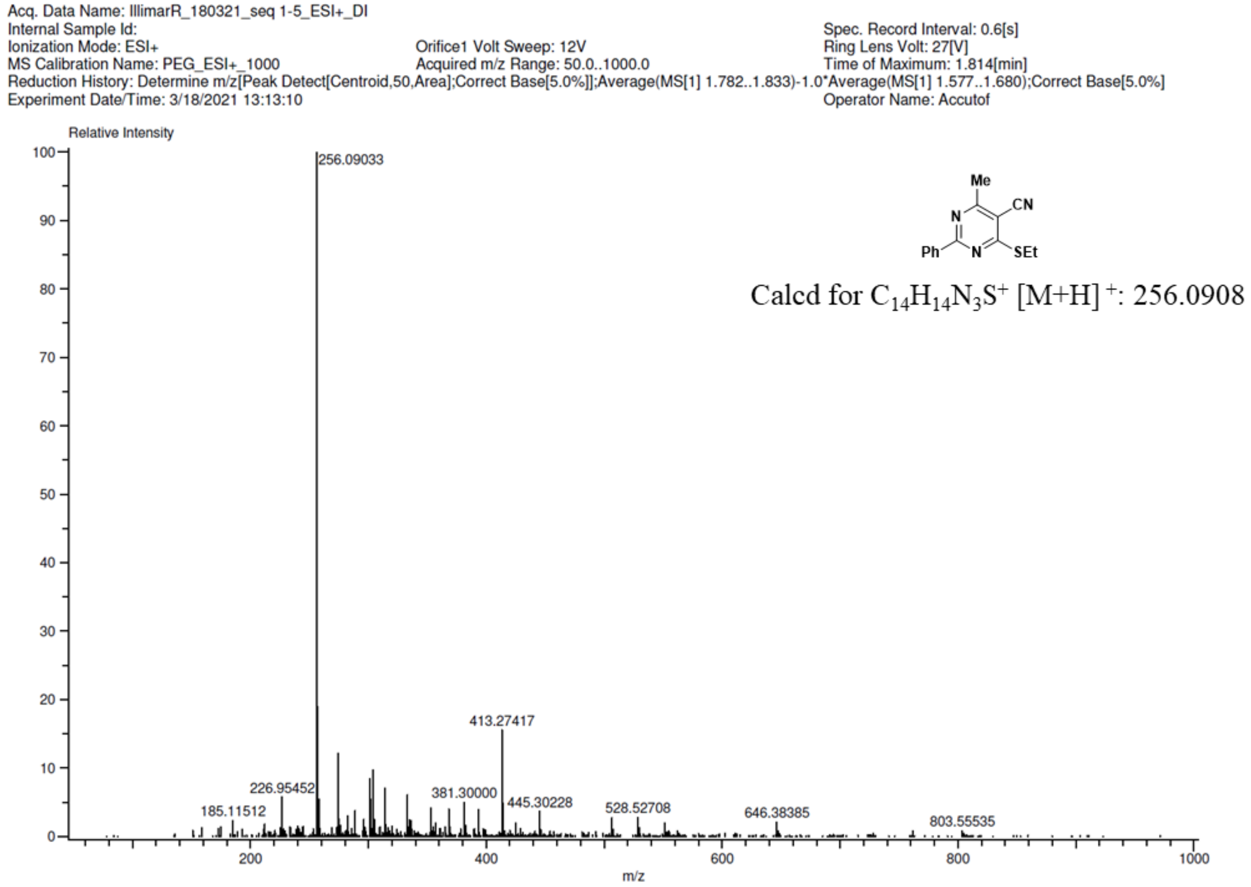
Spectrum 1: ^{1}H-NMR spectrum of compound ***6***



Spectrum 2: ^{13}C-NMR spectrum of compound ***6***



*Spectrum 3: HRMS spectrum of compound **6***

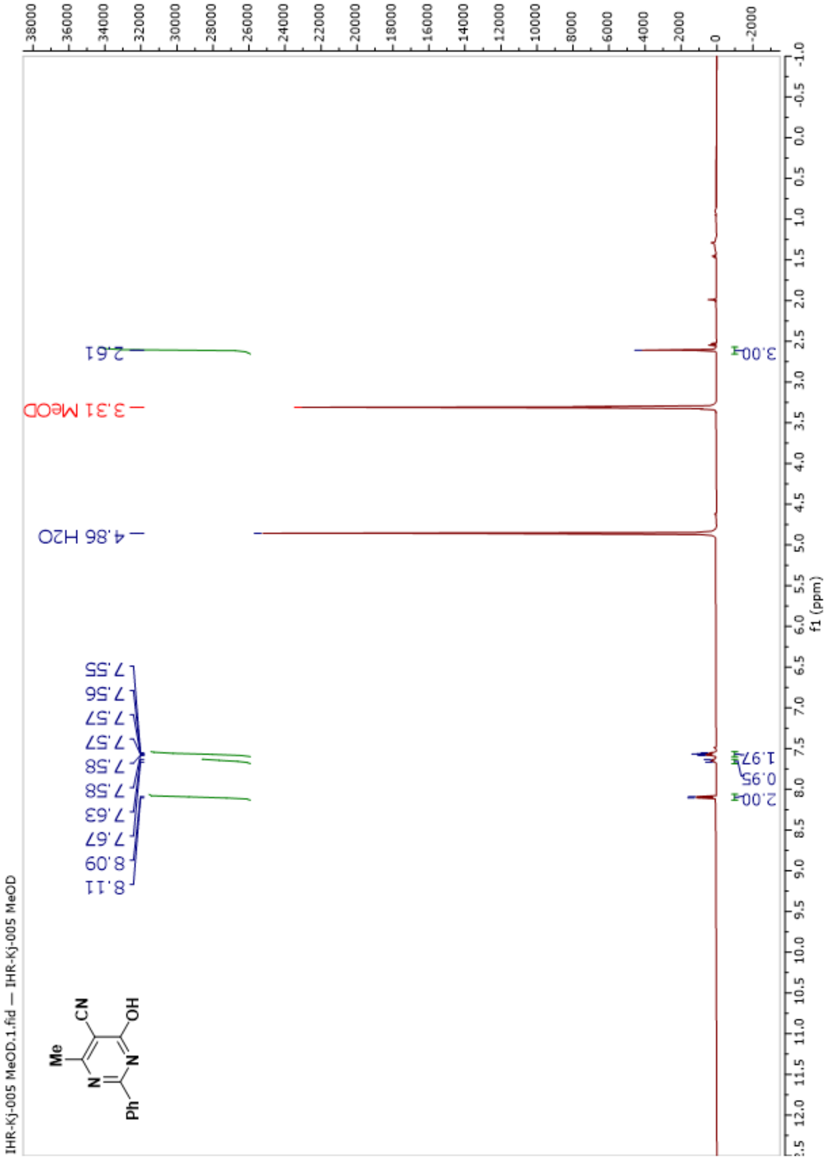


Acq. Data Name: IllimarR_180321_seq 1-5_ESI+_DI
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_1000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 1.782..1.833)-1.0*Average(MS[1] 1.577..1.680);Correct Base[5.0%]
Experiment Date/Time: 3/18/2021 13:13:10
Orifice1 Volt Sweep: 12V
Acquired m/z Range: 50.0..1000.0
Spec. Record Interval: 0.6[s]
Ring Lens Volt: 27[V]
Time of Maximum: 1.814[min]
Operator Name: Accutof

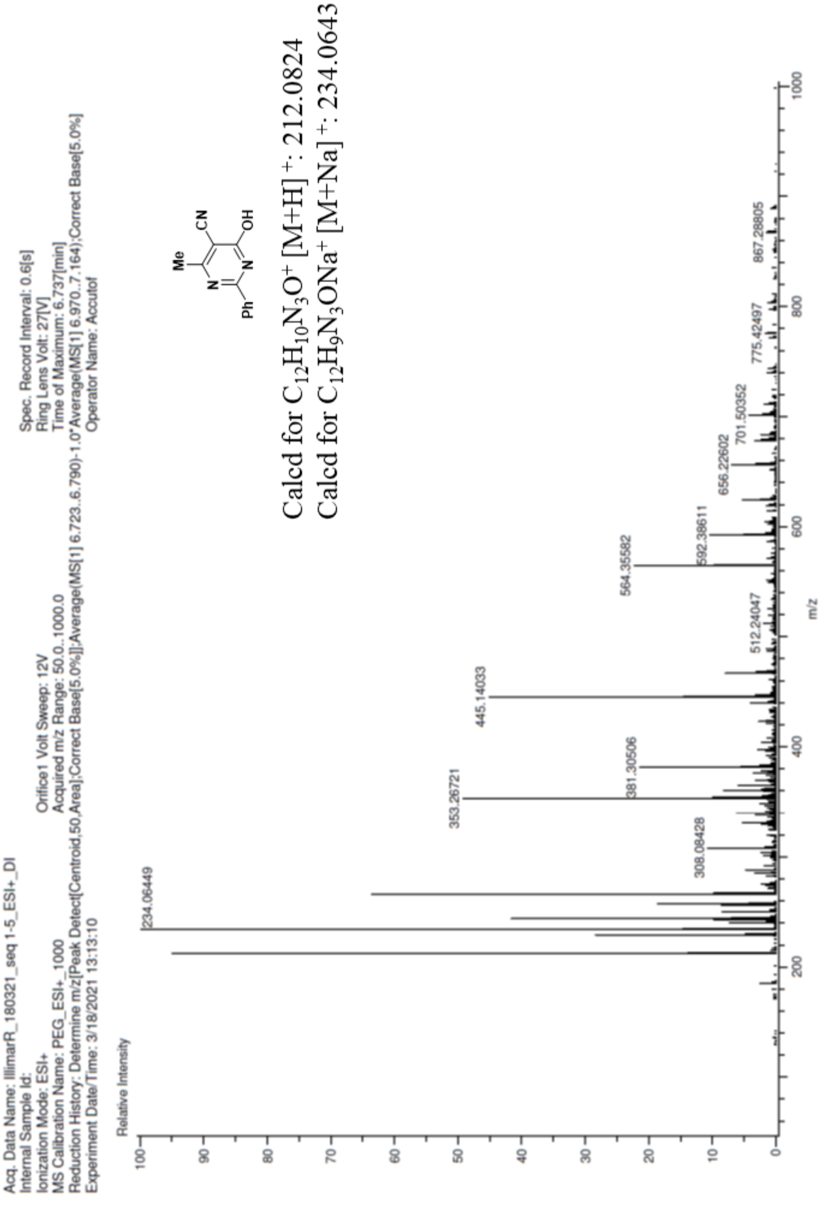
Calcd for $C_{14}H_{14}N_3S^+$ $[M+H]^+$: 256.0908

Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
256.09033
413.27417
226.95452
185.11512
381.30000
445.30228
528.52708
646.38385
803.55535
200
400
600
800
1000
m/z
Me
CN
N
N
Ph
SEt

Spectrum 4: ^{13}C-NMR spectrum of compound ***7***



Spectrum 5: HRMS spectrum of compound ***7***



Acq. Data Name: IllimarR_180321_seq 1-5_ESI+_DI
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_1000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 6.723..6.790)-1.0*Average(MS[1] 6.970..7.164);Correct Base[5.0%]
Experiment Date/Time: 3/18/2021 13:13:10

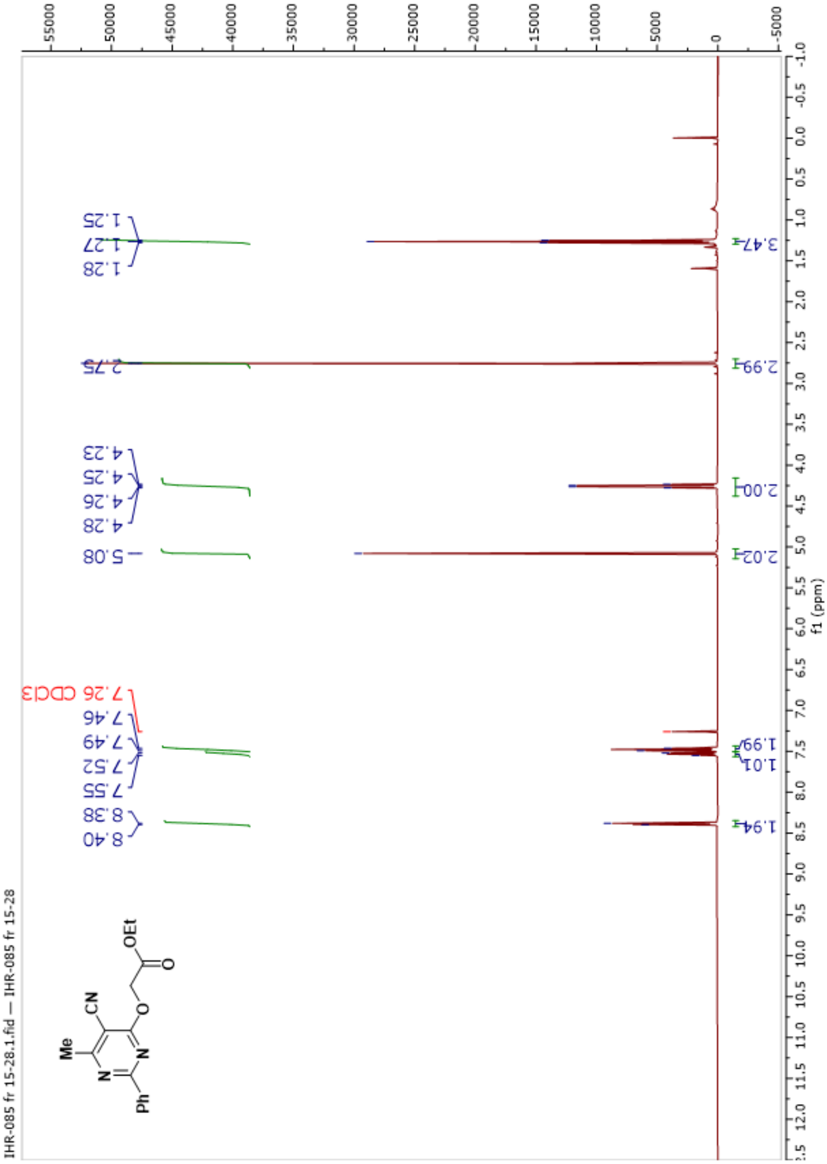
Orifice1 Volt Sweep: 12V
Acquired m/z Range: 50.0..1000.0

Spec. Record Interval: 0.6[s]
Ring Lens Volt: 27[V]
Time of Maximum: 6.737[min]
Operator Name: Accutof

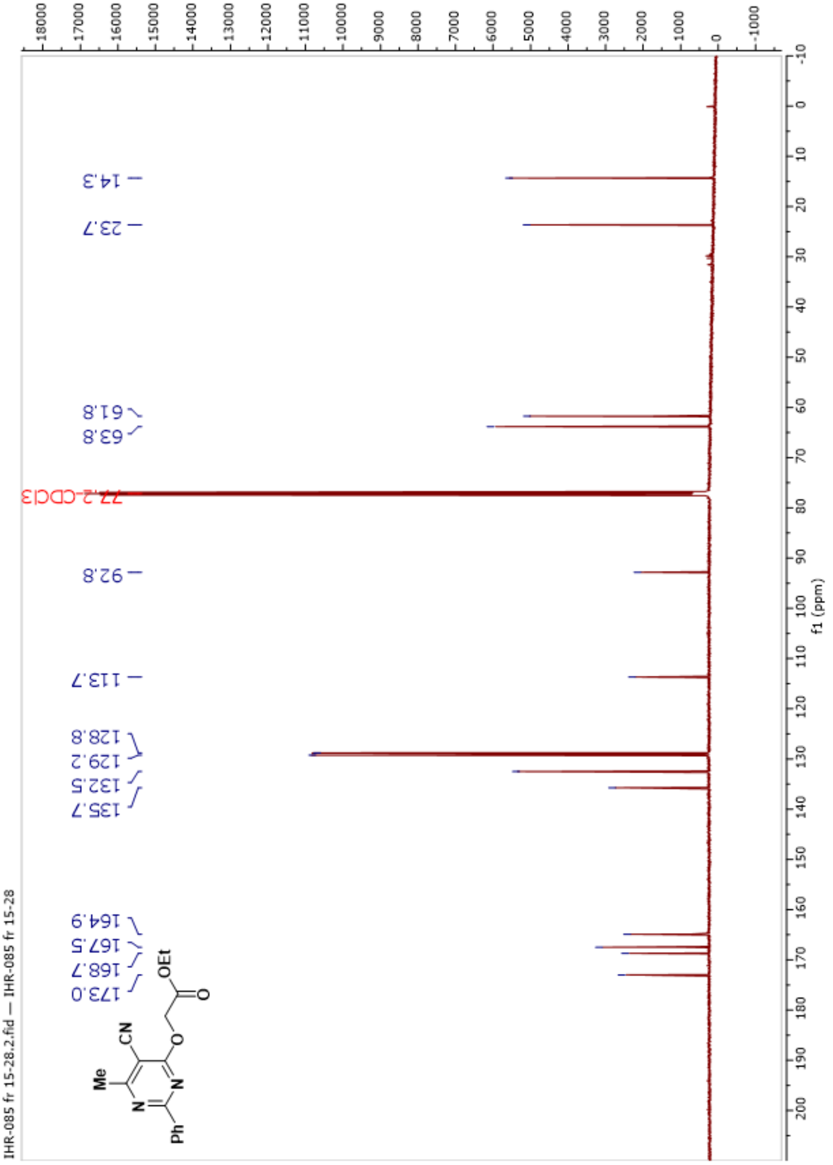
Calcd for $C_{12}H_{10}N_3O^+$ $[M+H]^+$: 212.0824
Calcd for $C_{12}H_9N_3ONa^+$ $[M+Na]^+$: 234.0643

Me
CN
N
N
Ph
OH
Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
200
400
600
800
1000
m/z
234.06449
308.08428
353.26721
381.30506
445.14033
512.24047
564.35582
592.38611
656.22602
701.50352
775.42497
867.28805

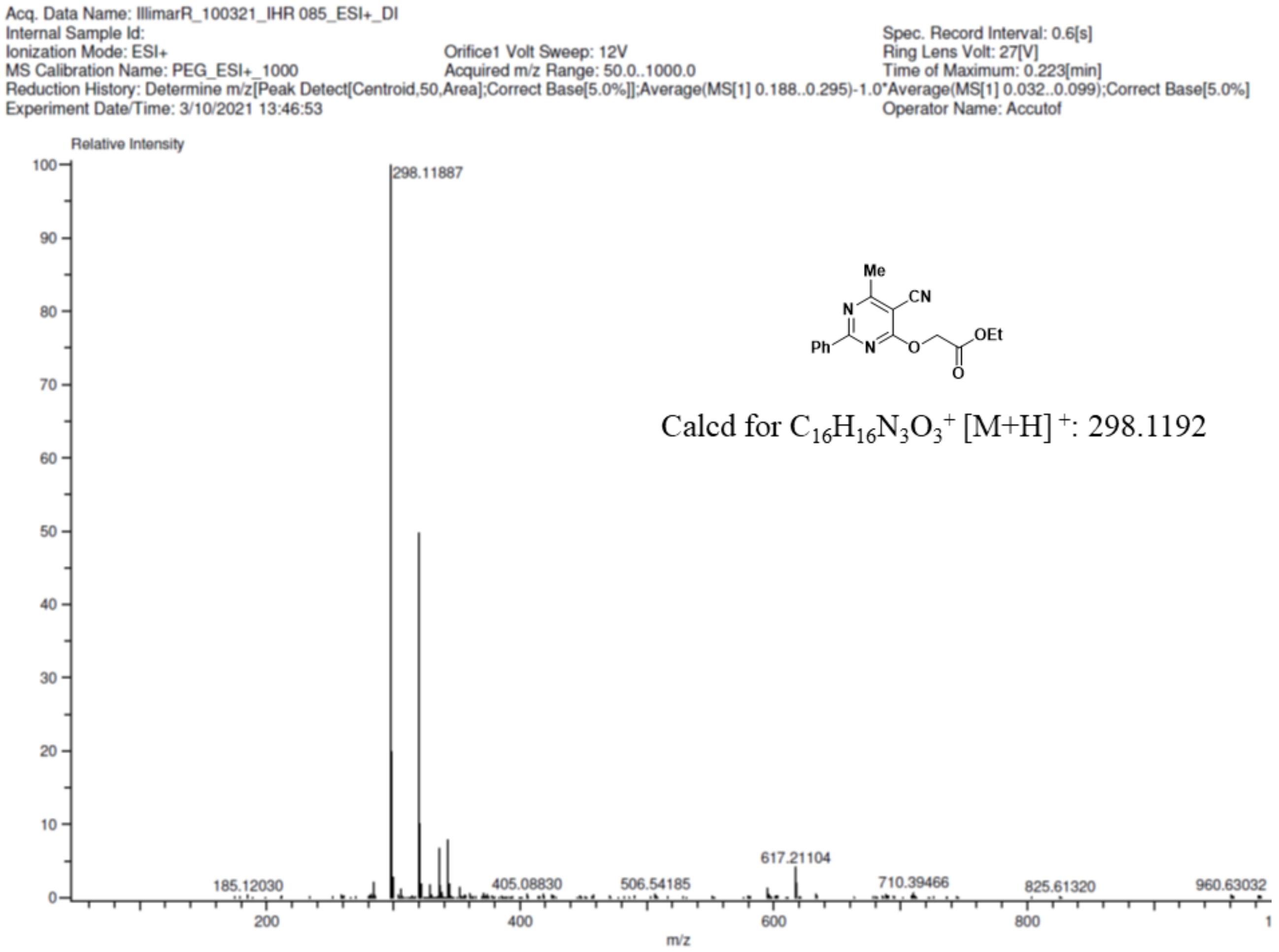
Spectrum 6: ^{1}H-NMR spectrum of compound ***2a***



Spectrum 7: 13C-NMR spectrum of compound ***2a***



Spectrum 8: HRMS spectrum of compound ***2a***



Acq. Data Name: IllimarR_100321_IHR 085_ESI+_DI
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_1000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 0.188..0.295)-1.0*Average(MS[1] 0.032..0.099);Correct Base[5.0%]
Experiment Date/Time: 3/10/2021 13:46:53

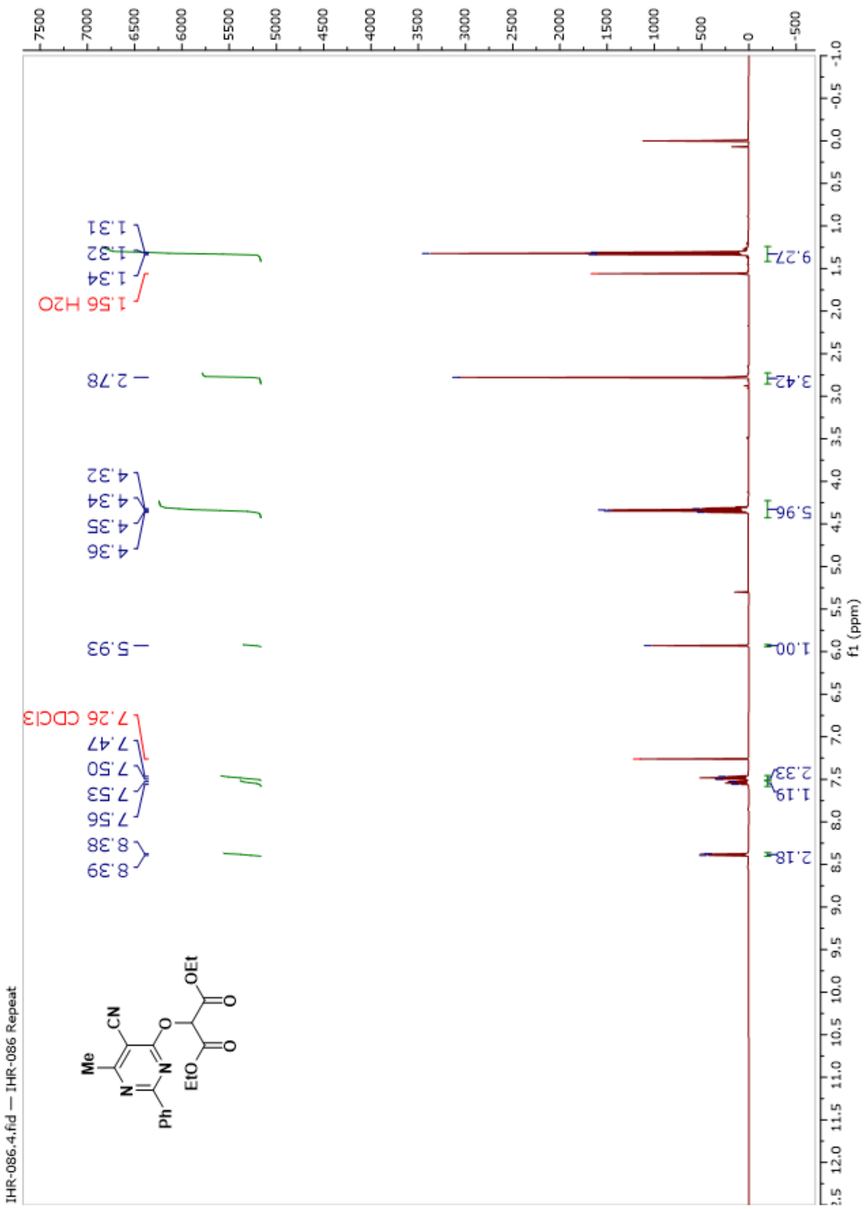
Orifice1 Volt Sweep: 12V
Acquired m/z Range: 50.0..1000.0

Spec. Record Interval: 0.6[s]
Ring Lens Volt: 27[V]
Time of Maximum: 0.223[min]
Operator Name: Accutof

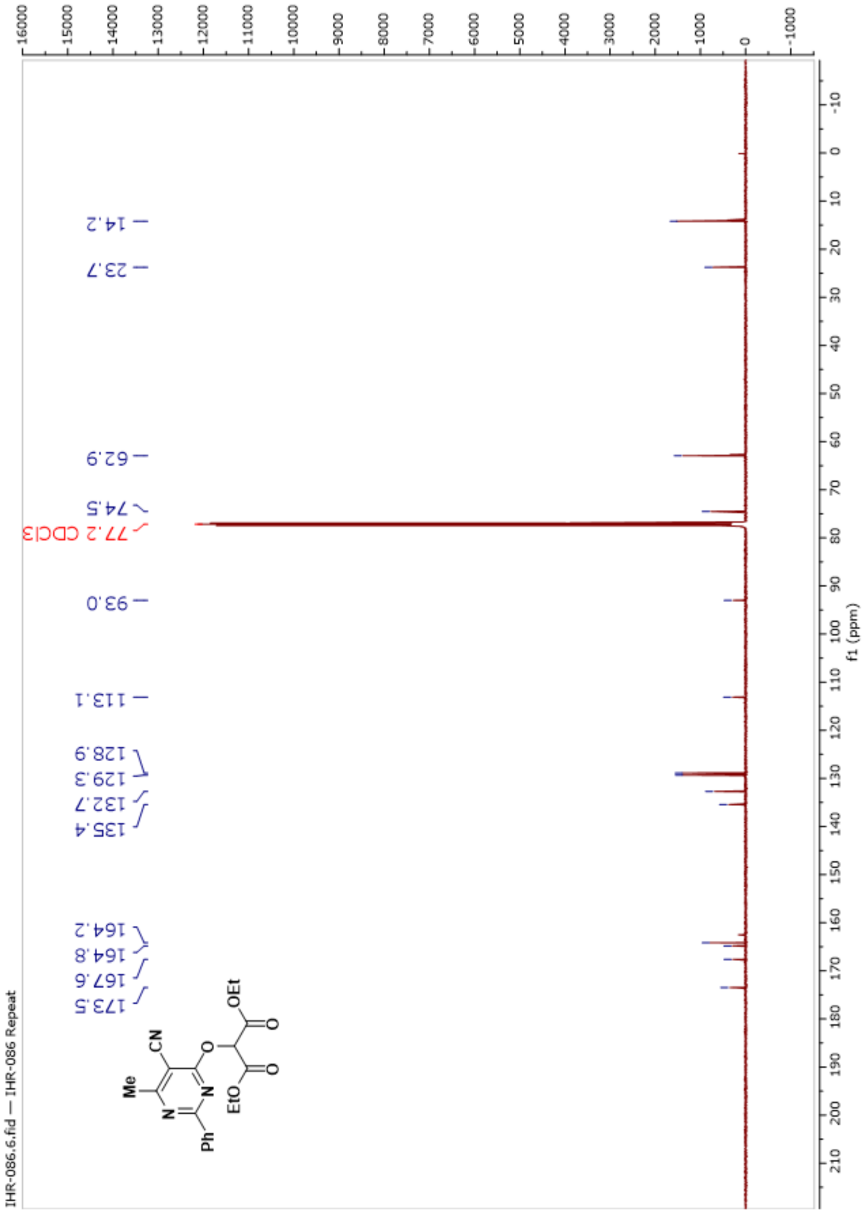
Calcd for $C_{16}H_{16}N_3O_3^+$ [M+H] $^+$: 298.1192

Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
298.11887
185.12030
405.08830
506.54185
617.21104
710.39466
825.61320
960.63032
200
400
600
800
1
m/z
Me
CN
N
N
Ph
O
OEt
O

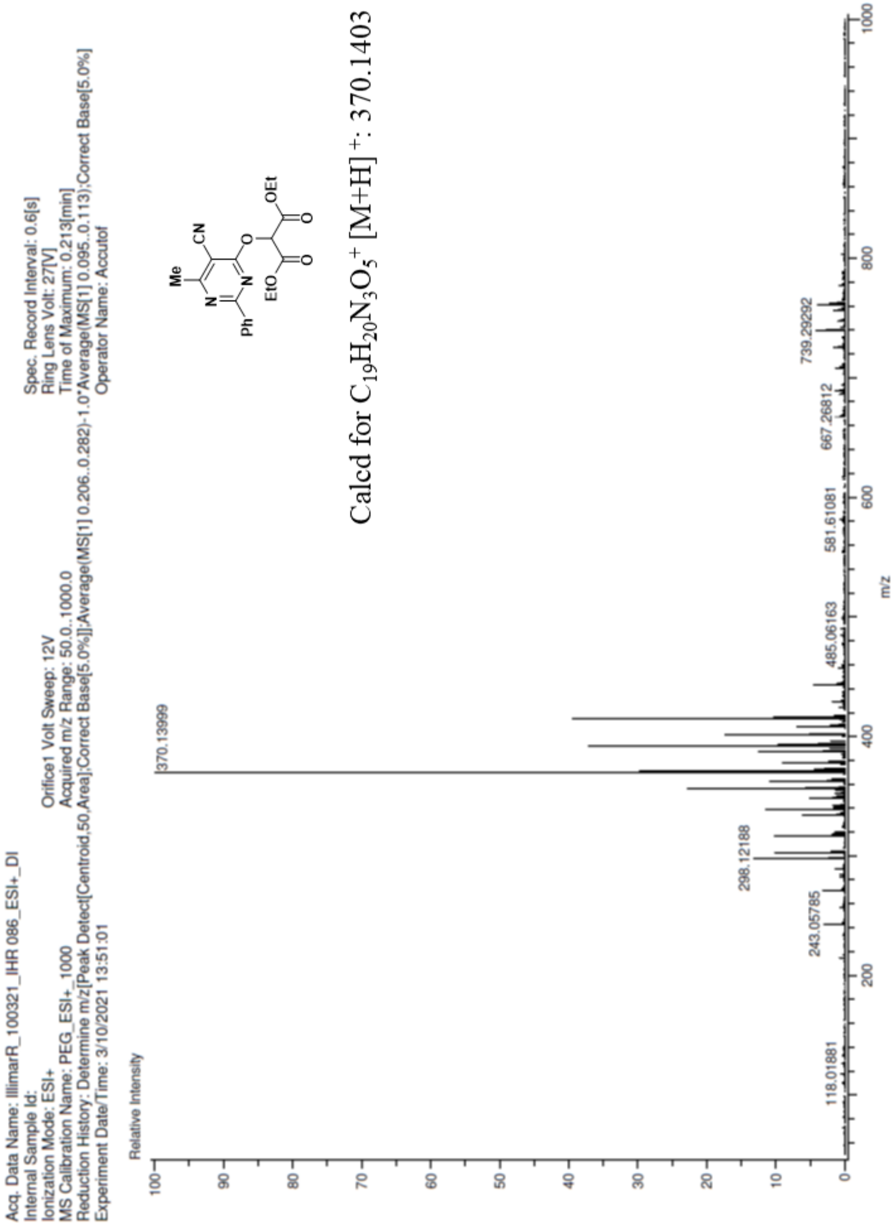
Spectrum 9: ^{1}H-NMR spectrum of compound ***2b***



Spectrum 10: ^{13}C-NMR spectrum of compound ***2b***



Spectrum 11: HRMS spectrum of compound ***2b***

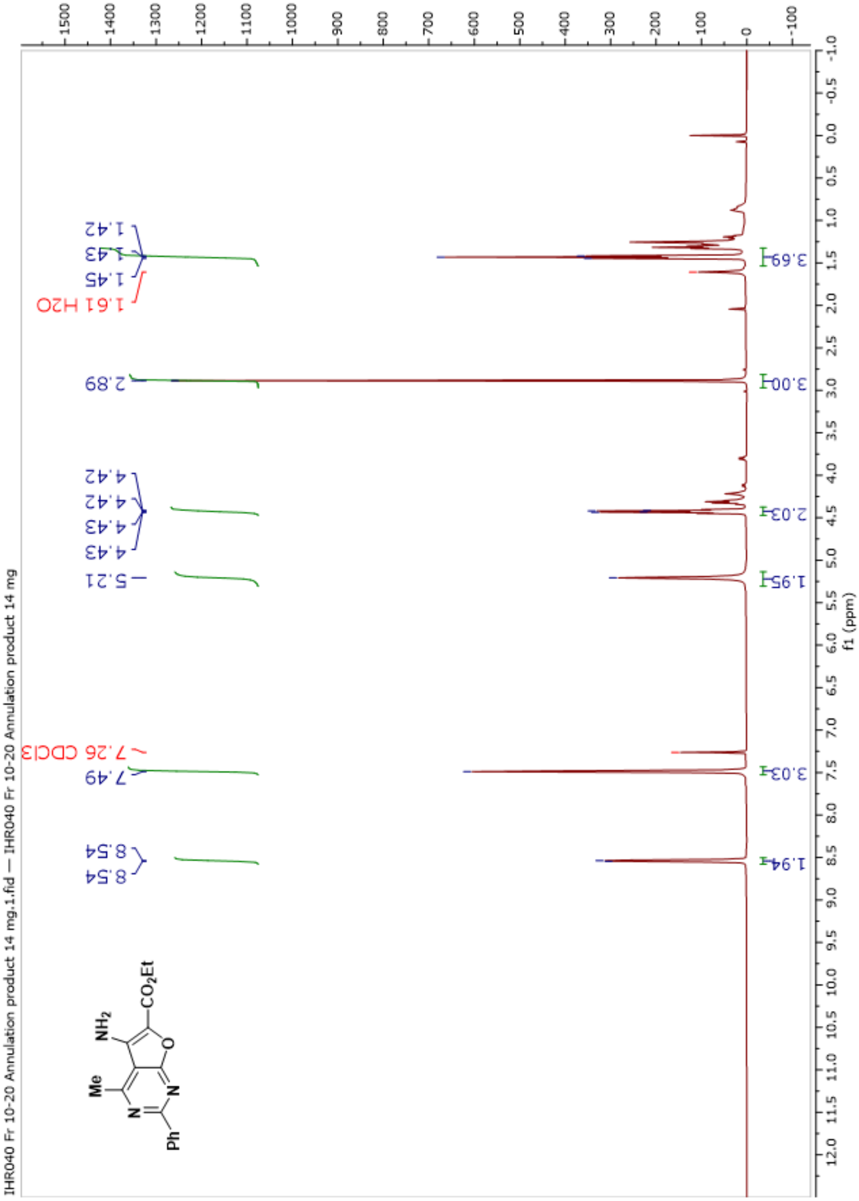


Acq. Data Name: IllimarR_100321_IHR 086_ESI+_DI
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_1000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 0.206..0.282)-1.0*Average(MS[1] 0.095..0.113);Correct Base[5.0%]
Experiment Date/Time: 3/10/2021 13:51:01
Orifice1 Volt Sweep: 12V
Acquired m/z Range: 50.0..1000.0
Spec. Record Interval: 0.6[s]
Ring Lens Volt: 27[V]
Time of Maximum: 0.213[min]
Operator Name: Accutof

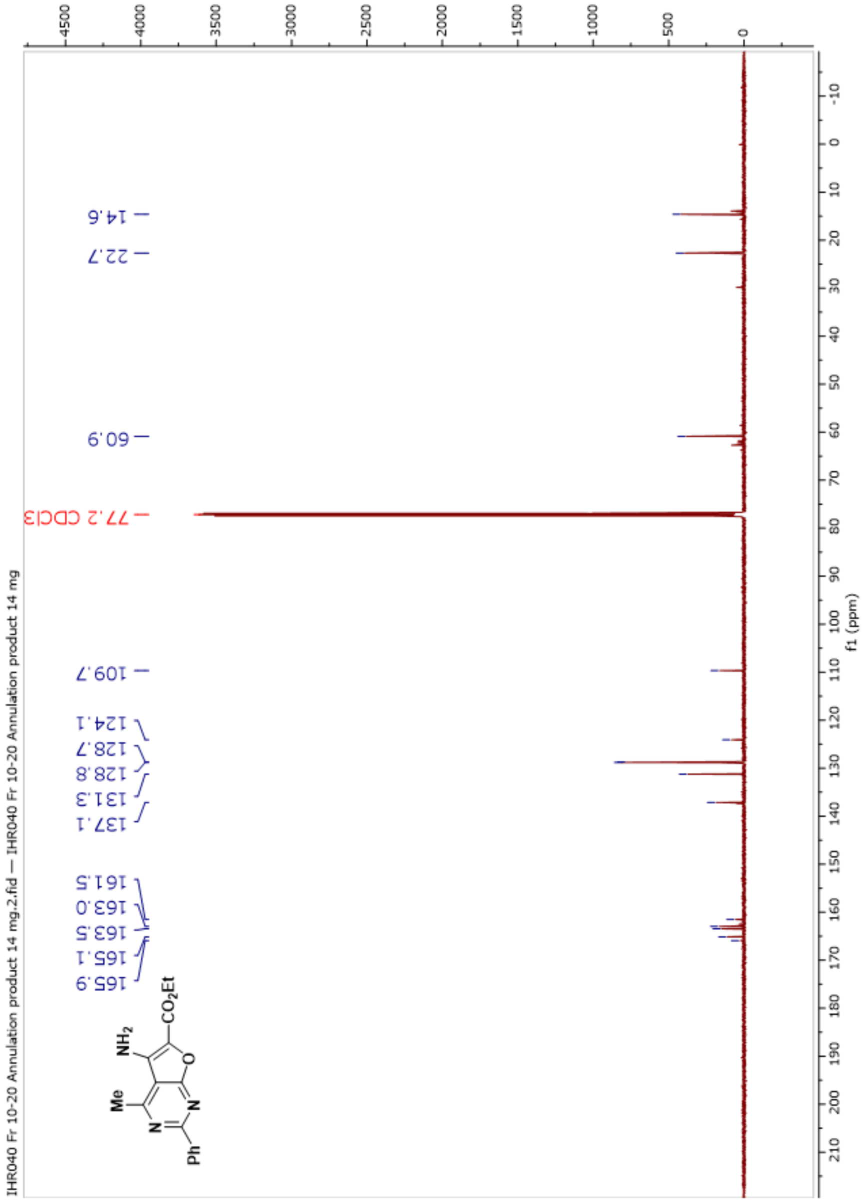
Calcd for $C_{19}H_{20}N_3O_5^+$ $[M+H]^+$: 370.1403

Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
200
400
600
800
1000
m/z
370.13999
118.01881
243.05785
298.12188
485.06163
581.61081
667.26812
739.29292
Me
CN
N
N
Ph
O
EtO
O
O
OEt

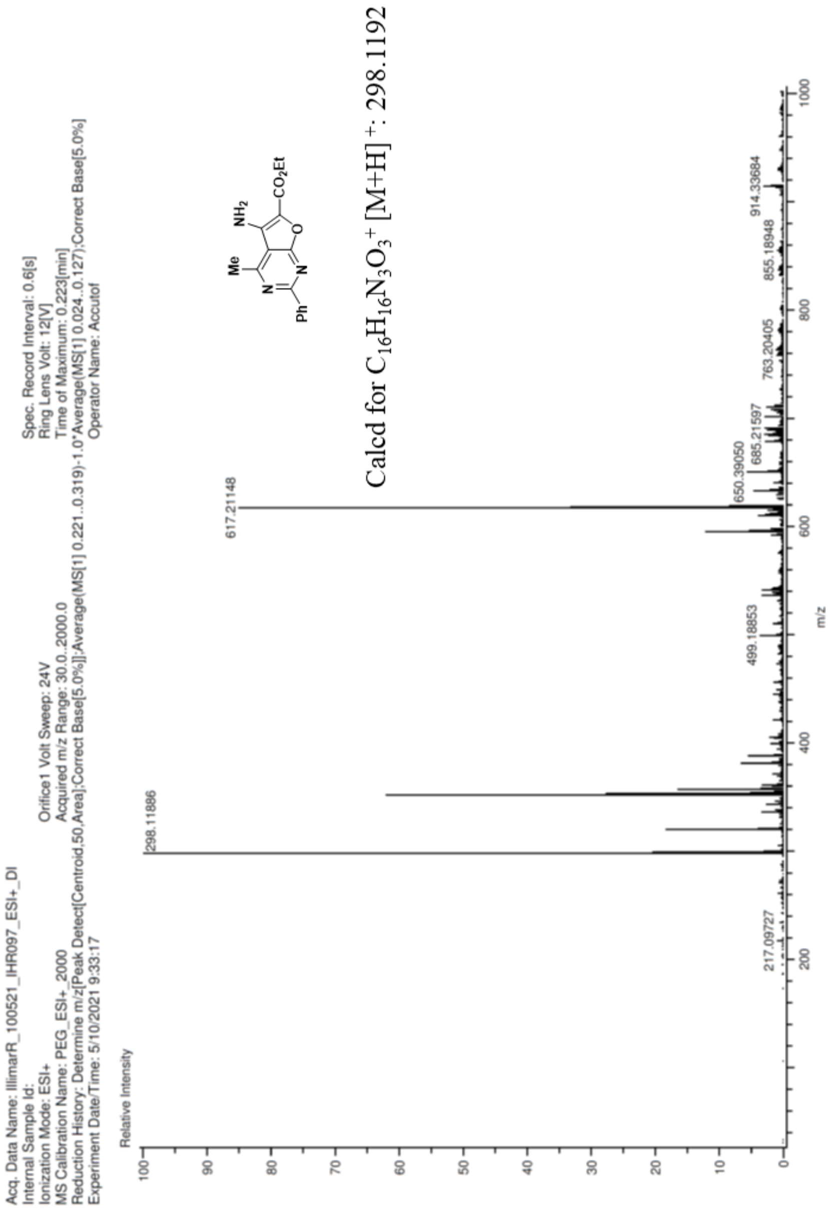
*Spectrum 12: ^{1}H-NMR spectrum of compound **1***



Spectrum 13: ^{13}C-NMR spectrum of compound ***1***



Spectrum 14: HRMS spectrum of compound ***1***

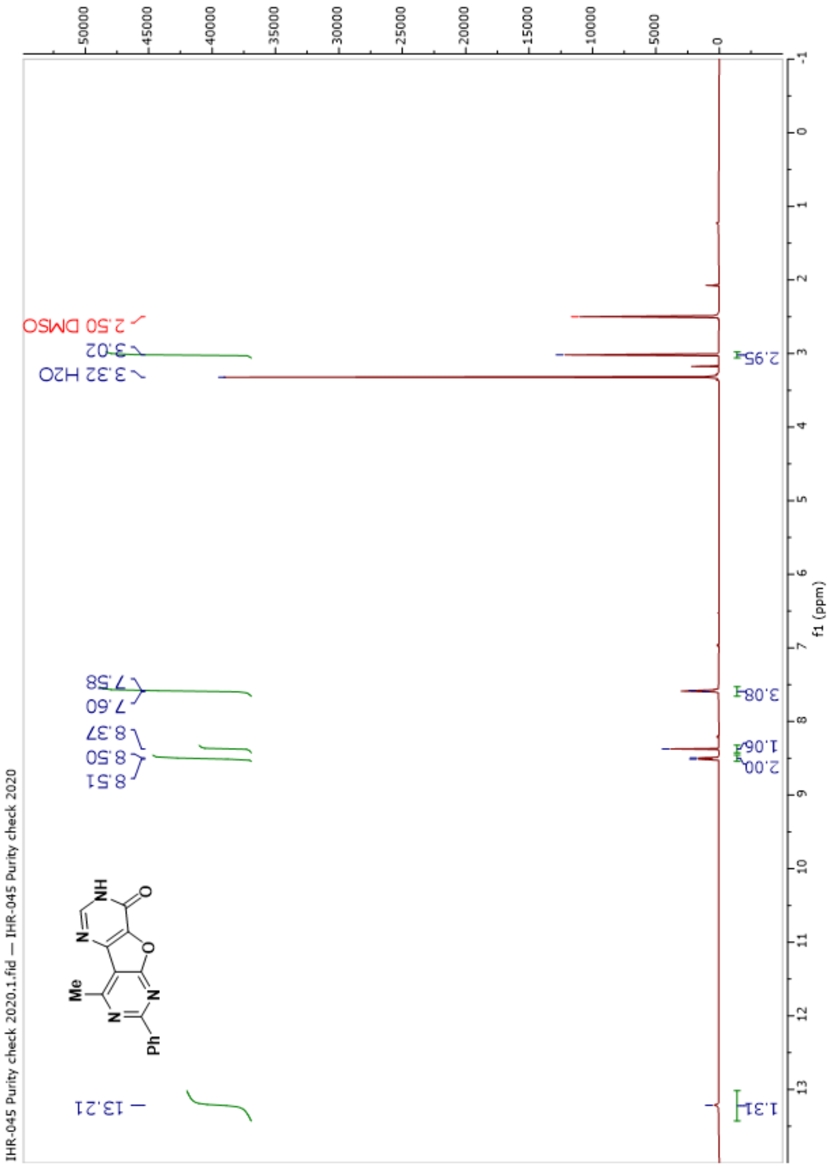


Acq. Data Name: IllimarR_100521_IHR097_ESI+_DI
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_2000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 0.221..0.319)-1.0*Average(MS[1] 0.024..0.127);Correct Base[5.0%]
Experiment Date/Time: 5/10/2021 9:33:17
Orifice1 Volt Sweep: 24V
Acquired m/z Range: 30.0..2000.0
Spec. Record Interval: 0.6[s]
Ring Lens Volt: 12[V]
Time of Maximum: 0.223[min]
Operator Name: Accutof

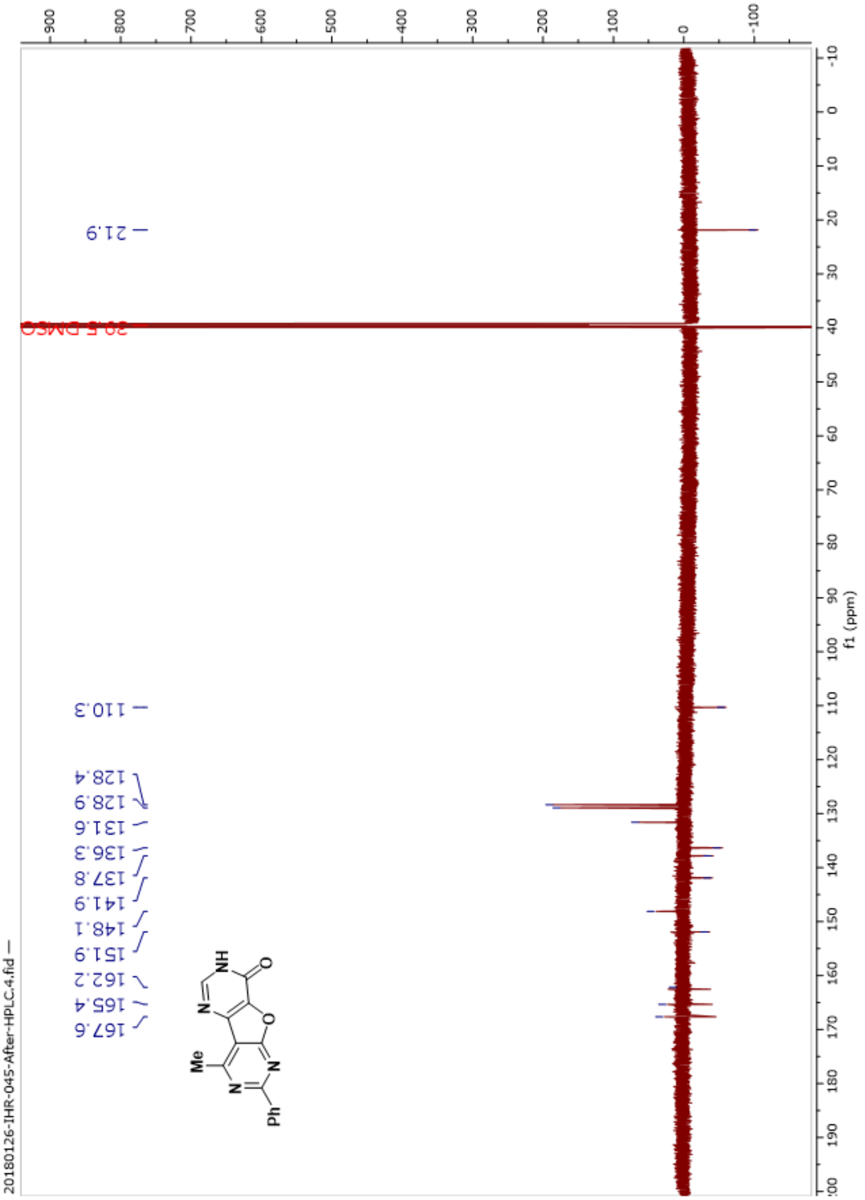
Calcd for $C_{16}H_{16}N_3O_3^+$ $[M+H]^+$: 298.1192

Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
200
400
600
800
1000
m/z
217.09727
298.11886
499.18853
617.21148
650.39050
685.21597
763.20405
855.18948
914.33684
Me
NH2
CO2Et
N
N
O
Ph

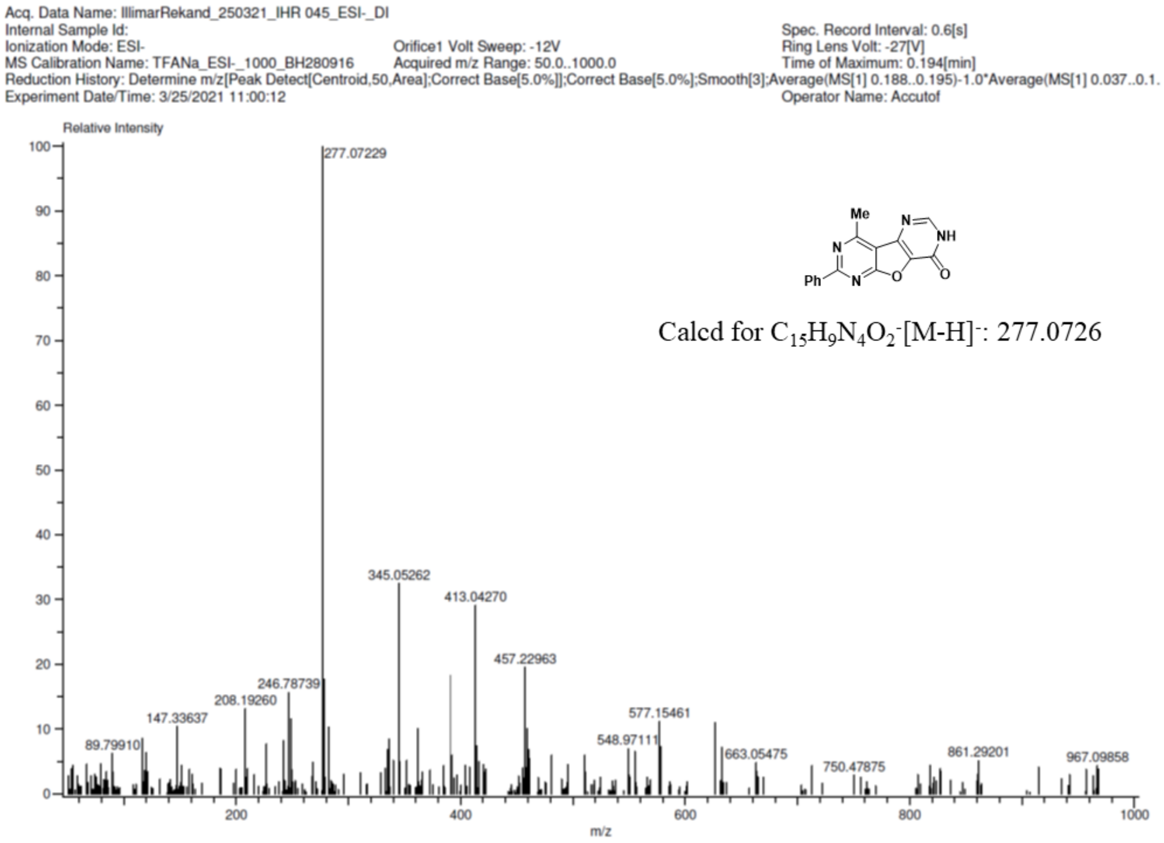
Spectrum 15: ^{1}H-NMR spectrum of ***RSL-0035***



Spectrum 16: ^{13}C-NMR (DEPT) spectrum of ***RSL-0035***



Spectrum 17: HRMS spectrum of ***RSL-0035***

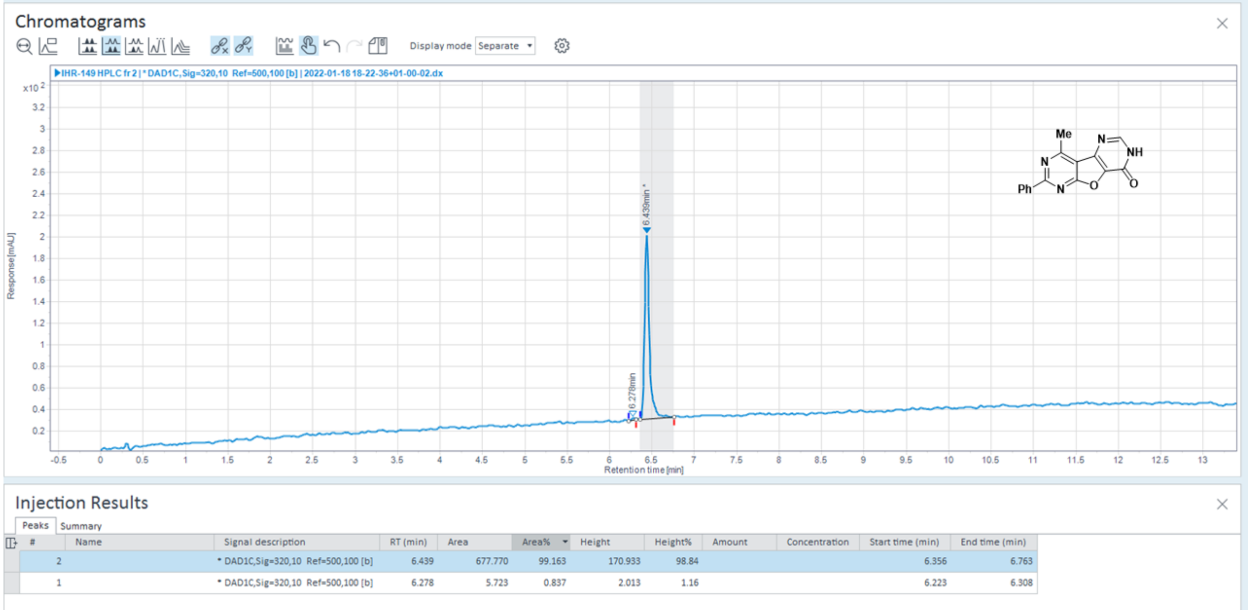


Acq. Data Name: IllimarRekand_250321_IHR 045_ESI-_DI
Internal Sample Id:
Ionization Mode: ESI-
MS Calibration Name: TFANa_ESI-_1000_BH280916
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Correct Base[5.0%];Smooth[3];Average(MS[1] 0.188..0.195)-1.0*Average(MS[1] 0.037..0.1.
Experiment Date/Time: 3/25/2021 11:00:12
Orifice1 Volt Sweep: -12V
Acquired m/z Range: 50.0..1000.0
Spec. Record Interval: 0.6[s]
Ring Lens Volt: -27[V]
Time of Maximum: 0.194[min]
Operator Name: Accutof

Calcd for $C_{15}H_9N_4O_2^-$[M-H]$^-$: 277.0726

Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
277.07229
345.05262
413.04270
457.22963
246.78739
208.19260
147.33637
89.79910
577.15461
548.97111
663.05475
750.47875
861.29201
967.09858
200
400
600
800
1000
m/z
Me
N
NH
N
Ph
N
O
O

Chromatogram 1: HPLC-chromatogram of RSL-0035


Chromatograms
Display mode Separate
▶IHR-149 HPLC fr 2 | * DAD1C,Sig=320,10 Ref=500,100 [b] | 2022-01-18 18-22-36+01-00-02.dx
x10 2
Response[mAU]
3.2
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5
9
9.5
10
10.5
11
11.5
12
12.5
13
Retention time [min]
6.439min *
6.278min
Me
N
NH
N
Ph
N
O
O
Injection Results
Peaks
Summary

#	Name	Signal description	RT (min)	Area	Area%	Height	Height%	Amount	Concentration	Start time (min)	End time (min)
2		* DAD1C,Sig=320,10 Ref=500,100 [b]	6.439	677.770	99.163	170.933	98.84			6.356	6.763
1		* DAD1C,Sig=320,10 Ref=500,100 [b]	6.278	5.723	0.837	2.013	1.16			6.223	6.308

6.2 Experiments leading up to and including **RSL-0031, RSL-0032, RSL-0033 and RSL-0034**

General procedure for tosylation:

The 3-aryl oxopropanenitrile (~5 mmol), *para*-toluenesulfonic anhydride (1.2 equivalents) were dissolved in anhydrous dichloromethane (25 mL). Triethyl amine (1.5 equivalents) was added dropwise over 15 minutes, and the solution was stirred at room temperature overnight, after which TLC analysis (Hx/ EtOAc, 6:1) indicated full conversion. The mixture was partitioned between dichloromethane and water (100 mL), and the aqueous phase was extracted with CH_2Cl_2 three times (3 x 100 mL). The combined organic layers were dried over $MgSO_4$, and solvent removed under reduced pressure to give a brown or black residue.

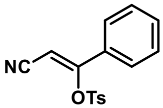
General procedure for amino malonate annulation:

The tosylate and diethyl amino malonate hydrochloride (1.2 equivalents) were dissolved in dry ethanol (5 mL) in a sealed microwave tube equipped with a stir bar. Freshly prepared sodium ethoxide from dry ethanol (1.6 M, 3.75 mL) was added over 15 min at 0 °C. The reaction mixture was heated to 80 °C and stirred overnight. The reaction mixture was partitioned between EtOAc (3 x 50 mL) and water, and the combined organic phases were dried over MgSO4. The crude product was purified on a Biotage Sfär silica gel cartridge (20 µm, 12 g) eluting with EtOAc in hexane (1 to 70% over 30 minutes, 40 mL/min), which provided the furane as a brown or black solid.

General procedure for annulation of amino esters using formamidine acetate

The amino ester and formamidine acetate (5 equivalents) were dissolved in absolute ethanol in a sealed microwave tube equipped with a stir bar, and heated to 80 °C. After 1 h the temperature was increased to 110 °C and continued stirred overnight. The crude material was purified using a Biotage Sfär C18 cartridge (20 µm, 50 g) eluting with acetonitrile in water, both containing 0.1% TFA (1 to 40% over 20 min, 15 ml/min to yield the pyrimidinones as an off-white or black solid.

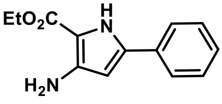
3-Phenyl-3-tosylacrylonitrile (**14a**)



The title compound was prepared from benzoyl acetonitrile (730 mg, 5.03 mmol) using the general tosylation conditions, and was purified on a silica gel column (Hx:EtOAc, 4:1). The pure *Z*-isomer was obtained as a beige solid after purification (0.87 g, 58%). The analytical data matched those previously published.[169]

^{1}H-NMR (CDCl$_3$, 500 MHz): δ = 2.47 (s, 3H), 5.56 (s, 3H), 7.35-7.50 (m, 5H), 7.58 (d, 2H, *J* = 8.0), 7.90 (d, 2H, *J* = 8.0)

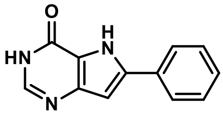
Ethyl 3-amino-5-phenylfuran-2-carboxylate (**15a**)



The title compound was prepared from tosylate **14a** (2.07 g, 6.94 mmol) using the general aminomalonate annulation conditions. The desired product eluted at ~50% EtOAc, and after removal of solvent isolated as a brown solid (1.13 g, 71%). The analytical data matched those previously published.[169]

^{1}H-NMR (500 MHz, (CD$_3$)$_2$SO): δ = 10.74 (s, 1H), 7.76-7.74 (m, 2H), 7.38 – 7.35 (m, 2H), 7.27 – 7.24 (m, 1H), 6.00 (d, *J* = 2.8, 1H), 5.10, (br s, 2H), 4.22 (q, *J* = 7.1 Hz, 2H), 1.30 (t, *J* = 7.1 Hz, 3H); ^{13}C-NMR (125 MHz, $(CD_3)_2SO$): δ = 161.1, 136.1, 131.4, 128.5, 127.3, 125.1, 105.9, 96.1, 80.3, 58.5, 14.8

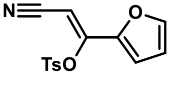
6-Phenyl-3,5-dihydro-4H-pyrrolo[3,2-d]pyrimidin-4-one (**RSL-0031**)



The title compound was prepared from pyrrole **15a** (49.7 mg, 0.216 mmol) using the general procedure for amino ester annulation. The reaction mixture was filtrated, and the precipitate was collected to give the product as a white solid (16.8 mg, 37%). The analytical data matched those previously published.[169]

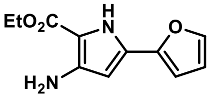
^{1}H-NMR (500 MHz, $(CD_3)_2SO$): δ = 12.31 (br s, 2H), 7.96 -7.94 (m, 2H), 7.81 (s, 1H), 7.44 (t, *J* = 7.7, 1H), 7.38-7.31 (m, 1H), 6.88 (s, 1H); ^{13}C-NMR (125 MHz, $(CD_3)_2SO$): δ = 153.5, 145.4, 142.1, 139.7, 131.2, 128.8, 128.1, 125.6, 119.0, 100.7; HRMS (ESI) *m/z* calcd for $C_{12}H_9N_3O^+$ [M+H]$^+$: 212.0824, found 212.0819

2-Cyano-1-(furan-2-yl)vinyl 4-methylbenzenesulfonate (**14b**)



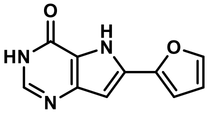
The title compound was prepared from 3-(furan-2-yl)-3-oxopropanenitrile (929 mg, 6.89 mmol) using the general tosylation method and isolated as a black solid (1.81 g, 91% crude yield). The crude material was taken directly into the next step without further purification.

Ethyl 3-amino-5-(furan-2-yl)-1H-pyrrole-2-carboxylate (**15b**)



The title compound was prepared from tosylate **14b** (1.37 g, 4.73 mmol) using the general amino malonate annulation conditions. The product was eluted at ~60% EtOAc and was isolated as an brown solid (0.687 g, 65% yield) after removal of solvent.

^{1}H-NMR (500 MHz, $(CD_3)_2SO$): δ = 10.87 (s, 1H), 7.64, (dd, *J* = 1.8, 0.8, 1H), 6.95 (dd, *J* = 3.4, 0.7, 1H), 6.53 (dd, *J* = 3.4, 1.8, 1H), 5.85 (d, *J* = 2.7, 1H), 5.12 (br. s, 2H), 4.22 (q, *J* = 7.1 Hz , 2H), 1.29 (t, *J* = 7.1 Hz, 3H); ^{13}C-NMR (125 MHz, $(CD_3)_2SO$): δ = 161.1, 147.1, 143.5, 142.2, 127.8, 111.7, 106.0, 105.2, 95.0, 58.6, 14.8; HRMS (ESI) *m/z* calcd for $C_{11}H_{13}N_2O_3^+$ $[M+H]^+$: 221.0926, found 221.0920

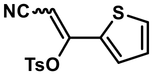


6-(Furan-2-yl)-3,5-dihydro-4H-pyrrolo[3,2-d]pyrimidin-4-one (**RSL-0032**)

The title compound was prepared from pyrrole **15b** (99 mg, 0.45 mmol) using the general amine ester annulation procedure. The product was eluted at ~30% acetonitrile and was isolated as a white solid (14.8 mg, 16%) after lyophilization.

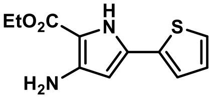
^{1}H-NMR (500 MHz, $(CD_3)_2SO$): δ = 12.51 (br s, 1H), 11.92, (br s, 1H), 7.81 (m, 1H), 7.77 (dd, *J* = 1.8, 0.88, 1H), 7.09 (dd, *J*= 3.4, 0.8, 1H), 6.62 (dd, *J* = 3.4, 1.8, 1H), 6.59 (s, 1H); ^{13}C-NMR (125 MHz, $(CD_3)_2SO$): δ = 153.4, 146.6, 145.1, 143.3, 142.2, 131.2, 118.6, 112.0, 107.3, 99.2; HRMS (ESI) *m/z* calcd For $C_{10}H_8N_3O_2^+$ [M+H]$^+$: 202.0617, found 202.0612

2-Cyano-1-(thiophen-2-yl)vinyl 4-methylbenzenesulfonate (**14c**)



The title compound was prepared from 3-oxo-3-(thiophen-2-yl)propanenitrile (1.040 g, 6.89 mmol) using the general tosylation method and isolated as a black solid (2.21 g, quantitative crude yield). The crude material was taken directly into the next step without further purification.

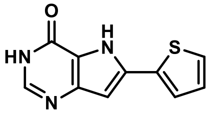
Ethyl 3-amino-5-(thiophen-2-yl)-1H-pyrrole-2-carboxylate (**15c**)



The title compound was prepared from tosylate **14c** (1.27 g, 4.16 mmol) using the general amino malonate annulation conditions. The product was eluted at ~60% ethyl acetate and isolated as a black solid (0.283 g, 24% yield) after removal of solvent.

^{1}H-NMR (500 MHz, $(CD_3)_2SO$): δ = 10.91 (s, 1H), 7.58 (dd, *J* = 3.4, 0.7, 1H), 7.44-7.43 (m, 1H), 7.06-7.05 (m, 1H), 5.79 (m, 1H), 5.12 (br. s, 2H), 4.22 (q, *J* = 7.2, 2H), 1.29 (t, *J* = 7.2 Hz, 3H); ^{13}C-NMR (125 MHz, $(CD_3)_2SO$) δ = 161.0, 144.2, 134.7, 130.7, 128.0, 124.8, 123.9, 105.3, 96.3, 58.5, 14.8; HRMS (ESI) *m/z* calcd for $C_{11}H_{13}N_2O_2S^+$ [M+H]$^+$: 237.0698, found 237.0692

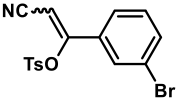
6-(Thiophen-2-yl)-3,5-dihydro-4H-pyrrolo[3,2-d]pyrimidin-4-one (**RSL-0033)**



The title compound was prepared from pyrrole **15c** (92 mg, 0.39 mmol) using the general amine ester annulation procedure. The product eluted at ~20% acetonitrile and was isolated as a black solid (14.8 mg, 16% yield) after lyophilization.

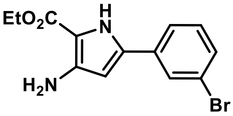
^{1}H-NMR (500 MHz, $(CD_3)_2SO$): δ = 12.56 (br s, 1H), 11.95 (br s, 1H), 7.84 (s, 1H) , 7.71 (dd, *J* = 3.6, 1.2, 1H), 7.57 (dd, *J* = 5.1, 1.2, 1H), 7. 13 (dd, *J* = 5.1, 3.6, 1H), 6.59 (d, *J* = 2.2 1H); ^{13}C-NMR (125 MHz, $(CD_3)_2SO$): δ = 153.3, 145.1, 142.3, 134.4, 134.1, 128.2, 126.4, 125.0, 118.6, 100.4; HRMS (ESI) *m/z* calcd for $C_{10}H_7N_3OS^+$ $[M+H]^+$: 218.0382, found 218.0383

1-(3-Bromophenyl)-2-cyanovinyl 4-methylbenzenesulfonate **(14d**)



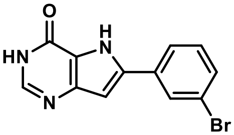
The title compound was prepared from 3-(3-bromophenyl)-3-oxopropanenitrile (1.00 g, 4.46 mmol) using the general tosylation method and isolated as a black solid (1.55 g, 92% crude yield). The crude material was taken directly into the next step without further purification.

Ethyl 3-amino-5-(3-bromophenyl)-1H-pyrrole-2-carboxylate (**15d**)



Tosylate **14d** (1.55 g, 4.11 mmol) and amino malonate hydrochloride (925 mg, 4.36 mmol) was dissolved in ethanol (20 mL). Potassium *tert*-butoxide (1.64 g, 14.6 mmol) dissolved in ethanol (10 mL) was added dropwise over 15 minutes. The reaction mixture was stirred overnight at room temperature and was extracted using EtOAc (3 x 100 mL) and water (100 mL). The combined organic layers were dried over $MgSO_4$. After removal of solvent, the crude product was isolated as a black tar (1.55 g, 178% crude yield, [assumed quantitative]) and used in the next step without further purification.

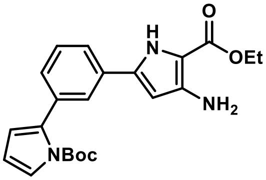
6-(3-Bromophenyl)-3,5-dihydro-4H-pyrrolo[3,2-d]pyrimidin-4-one (**16a**)



The title compound was prepared from pyrrole **15d** (642 mg, 2.08 mmol) using the general amine ester annulation procedure. The product was eluted at 28% acetonitrile and was isolated as a black solid (49.2 mg, 8% yield) after lyophilization.

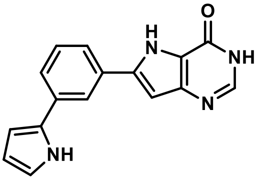
^{1}H-NMR (500 MHz, $(CD_3)_2SO$): δ = 12.60 (s, 1H), 12.09 (br s, 1H), 8.23 (t, *J* = 1.8, 1H), 7.97-7.95 (m, 1H), 7.93 (s, 1H), 7.54-7.52 (m, 1H), 7.40 (t, *J* = 8.4, 1H), 7.00 (d, *J* = 2.3, 1H); ^{13}C-NMR (125 MHz, $(CD_3)_2SO$)**:** δ = 153.3, 144.3, 142.5, 138.0, 133.3, 130.9, 130.8, 128.0, 124.6, 122.4, 119.4, 101.3; HRMS (ESI) *m/z* calcd for $C_{12}H_9BrN_3O^+$ [M+H]$^+$: 289.9929 found 289.9920

tert-Butyl 2-(3-(4-amino-5-(ethoxycarbonyl)-1H-pyrrol-2-yl)phenyl)-1H-pyrrole-1-carboxylate (**16b**)



Aryl bromide **15d** (493 mg, 1.60 mmol), Cs_2CO_3 (1.04 g, 3.20 mmol) and (1-(*tert*-butoxycarbonyl)-1H-pyrrol-2-yl)boronic acid (508 mg, 2.40 mmol) were dissolved in 1,4-dioxane (10 mL) in a microwave tube and the tube was sealed. The tube was flushed with argon and evacuated with a high vacuum pump, repeated three times. $Pd(dppf)_2Cl_2{\cdot}CH_2Cl_2$ (179 mg, 0.346 mmol) was added, and the resulting mixture was stirred while heated to 90 °C and stirred for 90 minutes. The resulting mixture was extracted with EtOAc (4 x 50 ml) and water (50 mL). The combined organic layers were dried over Na_2SO_4 and concentrated to give a solid material which was purified using a silica plug. The crude material was then further purified using a Biotage Sfär C18 cartridge (20 µm, 10 g) eluting with acetonitrile in water, both containing 0.1% TFA (40 to 70% over 20 min, 15 mL/min). The product was eluted at 56% acetonitrile and the title compound was isolated as a beige solid (140 mg, 20% yield) after lyophilization.

^{1}H-NMR (500 MHz, $(CD_3)_2SO$): δ = 11.98 (s, 1H), 7.93-7.92 (m, 1H), 7.74-7.73 (m, 2H), 7.42-7.28 (m, 2H), 7.28-7.26i (m, 1H), 6.53 (m, 1H), 6.34-6.33 (m, 1H), 6.31-6.30 (m, 1H), 4.29 (t, *J* = 7.1, 2H), 1.33 (t, *J* = 7.1, 3H), 1.27 (s, 9H); ^{13}C-NMR: Unable to determine all signals due to contamination; HRMS (ESI) *m/z* calcd for $C_{22}H_{26}N_3O_4^+$ [M+H]$^+$: 396.1923 found 396.1919

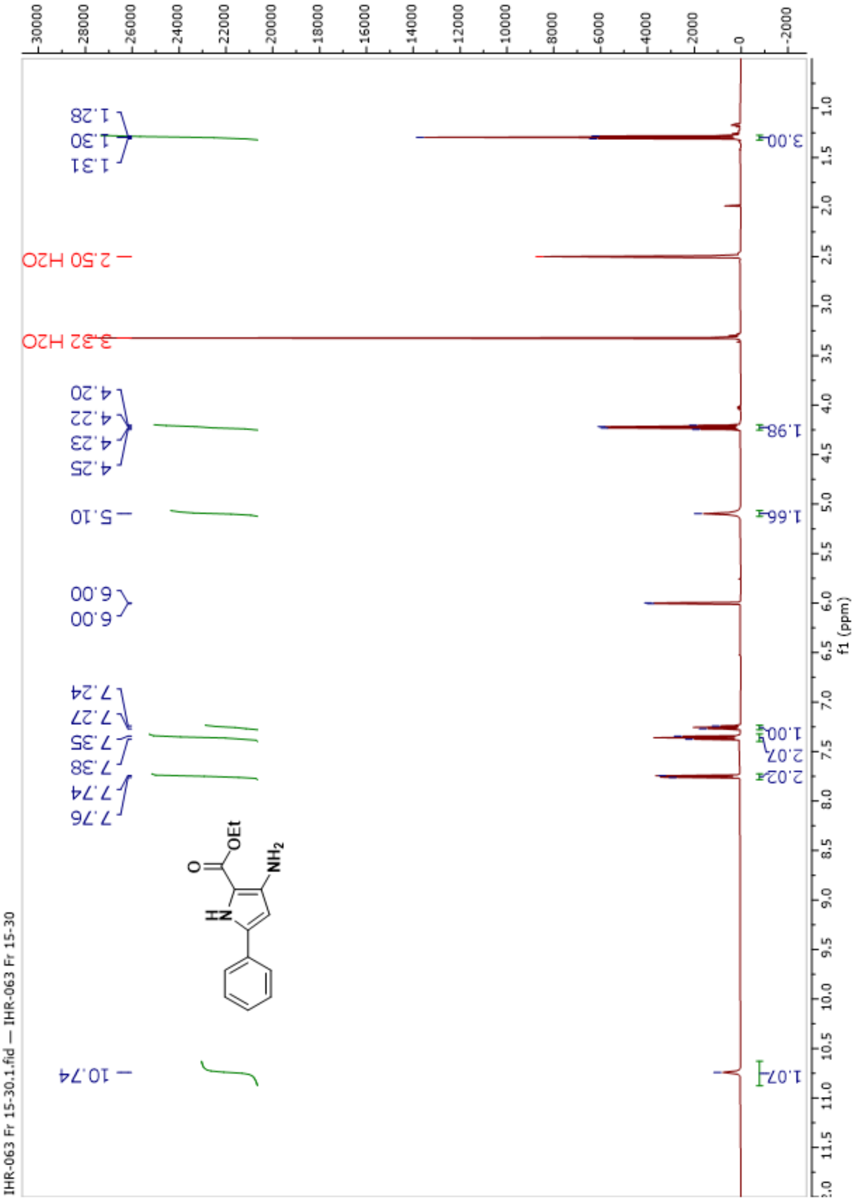


6-(3-(1H-pyrrol-2-yl)phenyl)-3,5-dihydro-4H-pyrrolo[3,2-d]pyrimidin-4-one (**RSL-0034**)

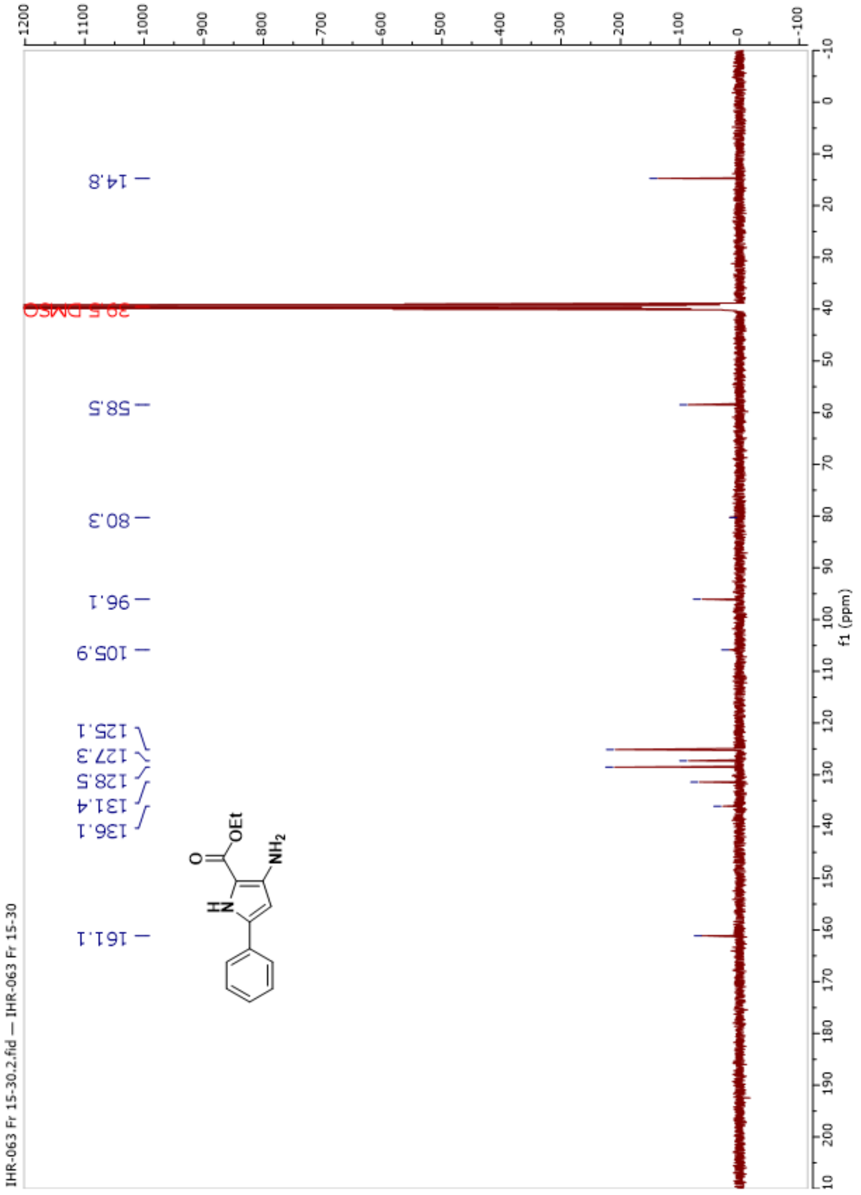
Pyrrole **11b** (21 mg, 0.053 mmol) was charged to a microwave tube with formamidinium acetate (41 mg, 0.39 mmol) and ethanol (3 mL). The solution was heated to 100 °C and stirred overnight. The resulting mixture was partitioned between EtOAc (3 x 50 mL) and water (50 mL) and the combined organic layers dried over $MgSO_4$ before concentrated *in vacuo*. The residue was then further purified using a Biotage Sfär C18 cartridge (20 µm, 10 g) eluting with acetonitrile in water, both containing 0.1% TFA (10 to 50% over 20 min, 15 mL/min). The product eluted at 38% acetonitrile and the title compound was isolated as a black solid (1.4 mg, 10% yield) after lyophilization.

^{1}H-NMR (600 MHz, $(CD_3)_2SO$): δ = 12.39 (s, 1H), 11.91 (br s, 1H), 11.27 (s, 1H), 8.24-8.24 (m, 1H), 7.84 (s, 1H), 7.73-7.71 (m, 1H), 7.59-7.57 (m, 1H), 7.40 (t, 1H, *J* = 7.7 Hz), 6.95 (d, *J* = 2.3, 1H), 6.90-6.89 (m, 1H), 6.69-6.68 (m, 1H), 6.16-6.15 (m, 1H); ^{13}C-NMR (151 MHz, $(CD_3)_2SO$): δ = 153.4, 142.1, 139.8, 133.5, 131.4, 130.8, 129.2, 122.7, 122.4, 120.5, 120.1, 119.5, 119.0, 109.1, 106.2, 100.8; HRMS (ESI) *m/z* calcd For $C_{16}H_{13}N_4O_1^+$ $[M+H]^+$: 277.1089, found 277.1085

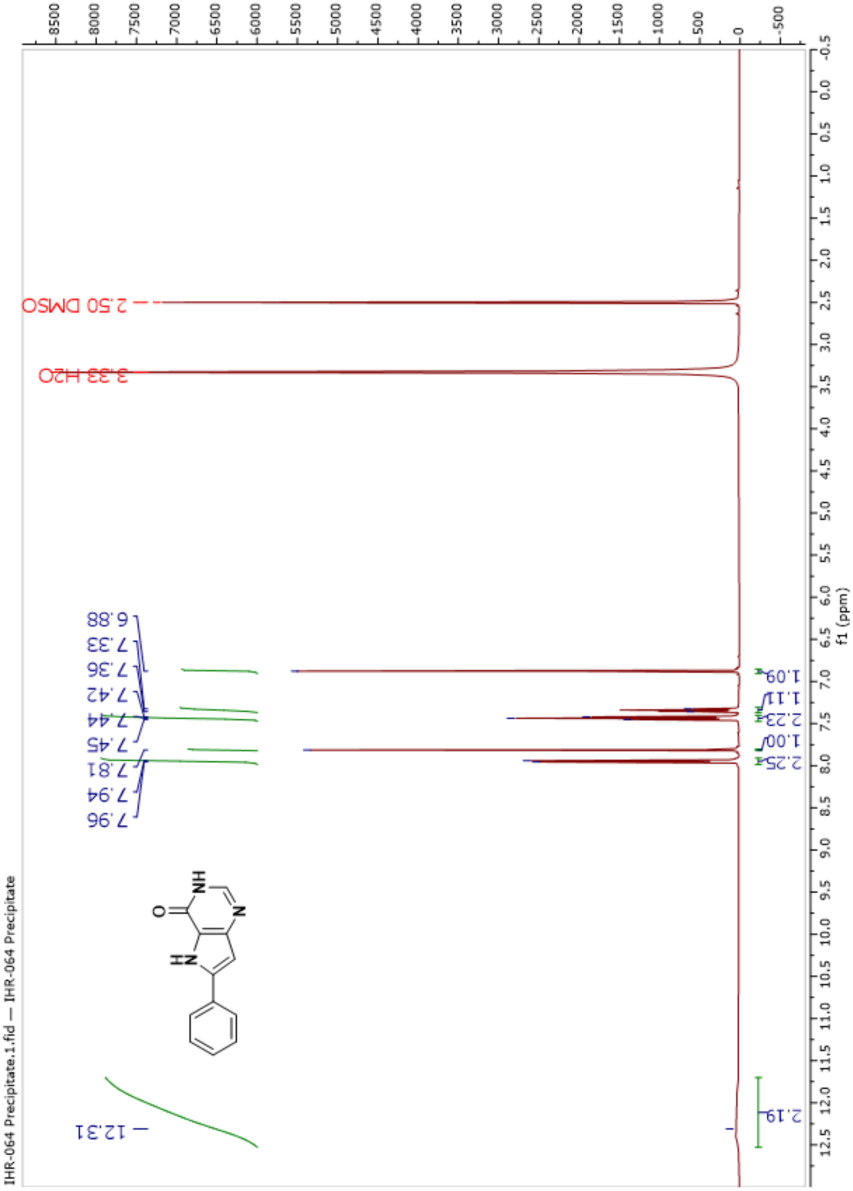
*Spectrum 18: ^{1}H NMR of compound **15a***



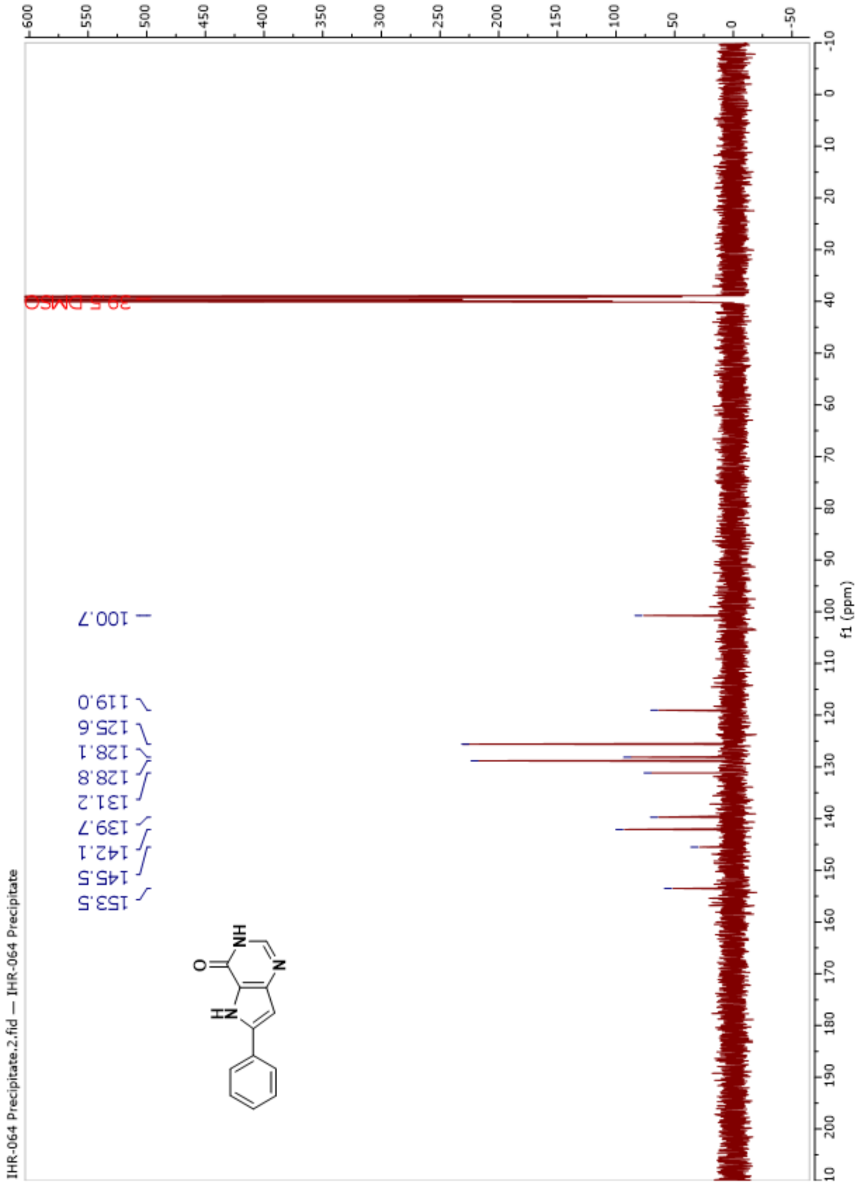
Spectrum 19: ^{13}C NMR of compound 15a



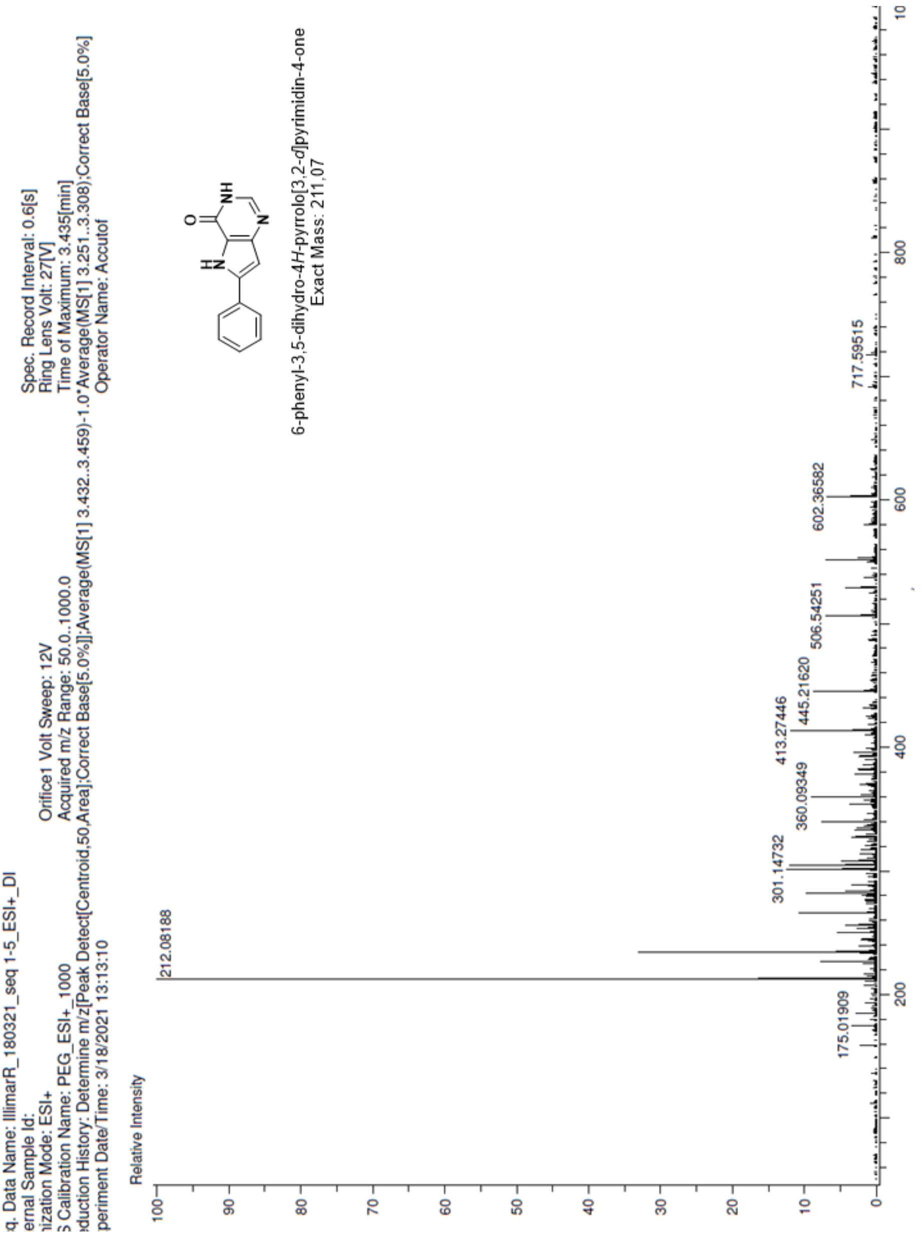
Spectrum 20: ^{1}H NMR of ***RSL-0031***



Spectrum 21: 13C NMR of ***RSL-0031***



Spectrum 22: HRMS of ***RSL-0031***

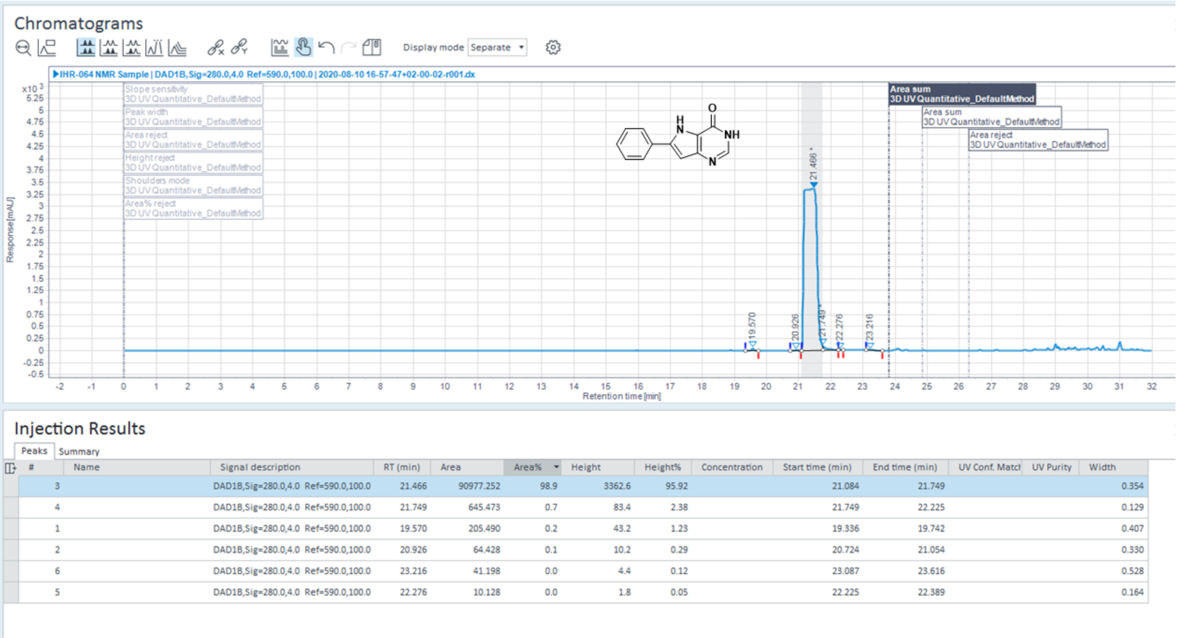


q. Data Name: IllimarR_180321_seq 1-5_ESI+_DI
ernal Sample Id:
nization Mode: ESI+
S Calibration Name: PEG_ESI+_1000
duction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 3.432..3.459)-1.0*Average(MS[1] 3.251..3.308);Correct Base[5.0%]
periment Date/Time: 3/18/2021 13:13:10
Orifice1 Volt Sweep: 12V
Acquired m/z Range: 50.0..1000.0
Spec. Record Interval: 0.6[s]
Ring Lens Volt: 27[V]
Time of Maximum: 3.435[min]
Operator Name: Accutof

6-phenyl-3,5-dihydro-4*H*-pyrrolo[3,2-*d*]pyrimidin-4-one
Exact Mass: 211,07

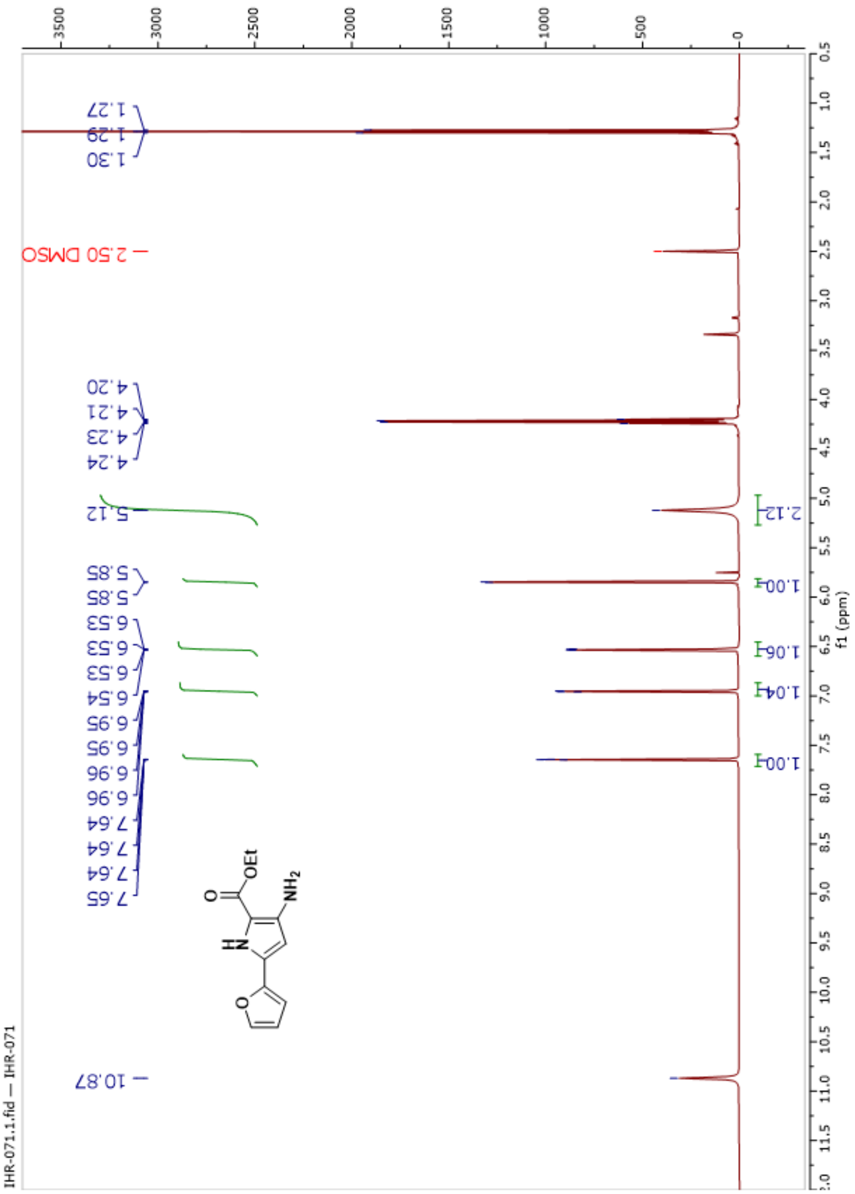
Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
212.08188
175.01909
301.14732
360.09349
413.27446
445.21620
506.54251
602.36582
717.59515
200
400
600
800
10
O
NH
N
H
N

Chromatogram 2: HPLC chromatogram of ***RSL-0031***

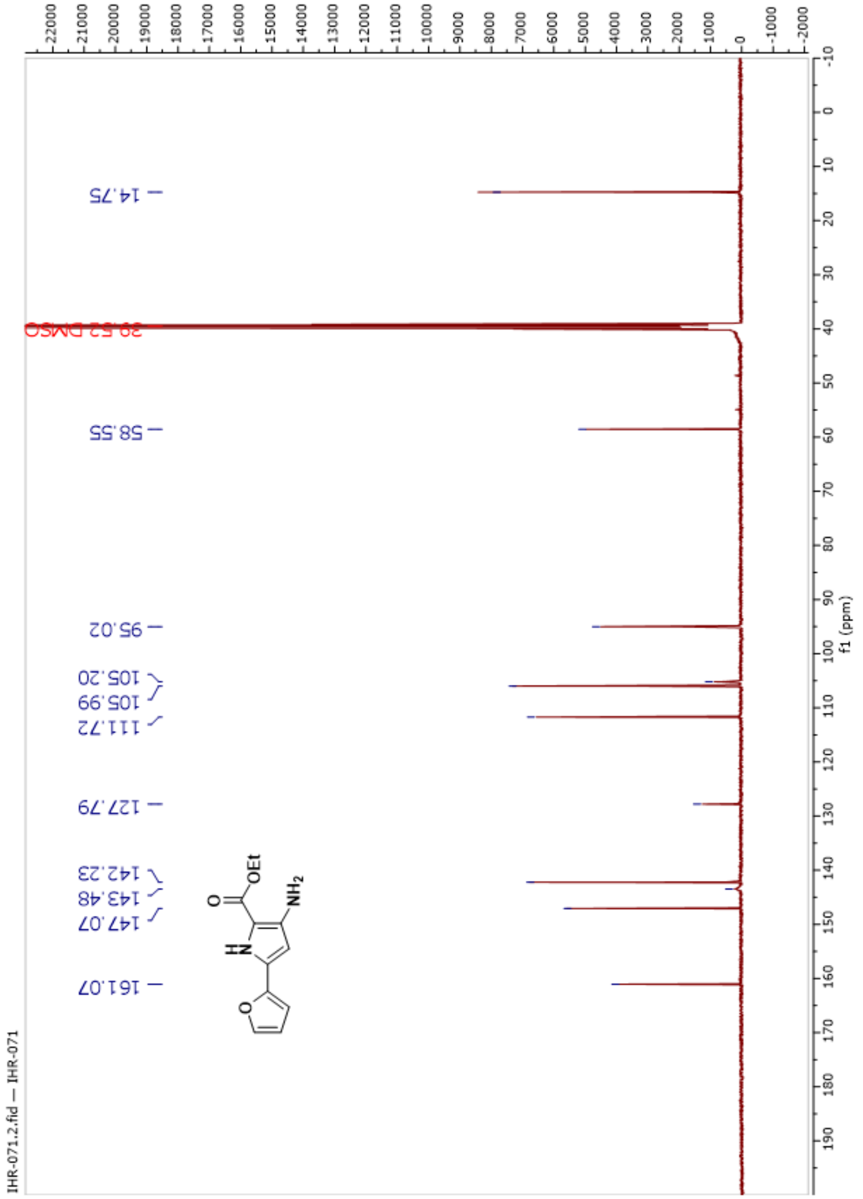

Chromatograms
Display mode Separate
▶IHR-064 NMR Sample | DAD1B,Sig=280.0,4.0 Ref=590.0,100.0 | 2020-08-10 16-57-47+02-00-02-r001.dx
Slope sensitivity 3D UV Quantitative_DefaultMethod
Peak width 3D UV Quantitative_DefaultMethod
Area reject 3D UV Quantitative_DefaultMethod
Height reject 3D UV Quantitative_DefaultMethod
Shoulders mode 3D UV Quantitative_DefaultMethod
Area% reject 3D UV Quantitative_DefaultMethod
Area sum 3D UV Quantitative_DefaultMethod
Area sum 3D UV Quantitative_DefaultMethod
Area reject 3D UV Quantitative_DefaultMethod
x10 3
Response [mAU]
5.25
5
4.75
4.5
4.25
4
3.75
3.5
3.25
3
2.75
2.5
2.25
2
1.75
1.5
1.25
1
0.75
0.5
0.25
0
-0.25
-0.5
-2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
Retention time [min]
19.570
20.926
21.466 *
21.749
22.276
23.216
H N
O
NH
N
Injection Results
Peaks
Summary

#	Name	Signal description	RT (min)	Area	Area%	Height	Height%	Concentration	Start time (min)	End time (min)	UV Conf. Matcl	UV Purity	Width
3		DAD1B,Sig=280.0,4.0 Ref=590.0,100.0	21.466	90977.252	98.9	3362.6	95.92		21.084	21.749			0.354
4		DAD1B,Sig=280.0,4.0 Ref=590.0,100.0	21.749	645.473	0.7	83.4	2.38		21.749	22.225			0.129
1		DAD1B,Sig=280.0,4.0 Ref=590.0,100.0	19.570	205.490	0.2	43.2	1.23		19.336	19.742			0.407
2		DAD1B,Sig=280.0,4.0 Ref=590.0,100.0	20.926	64.428	0.1	10.2	0.29		20.724	21.054			0.330
6		DAD1B,Sig=280.0,4.0 Ref=590.0,100.0	23.216	41.198	0.0	4.4	0.12		23.087	23.616			0.528
5		DAD1B,Sig=280.0,4.0 Ref=590.0,100.0	22.276	10.128	0.0	1.8	0.05		22.225	22.389			0.164

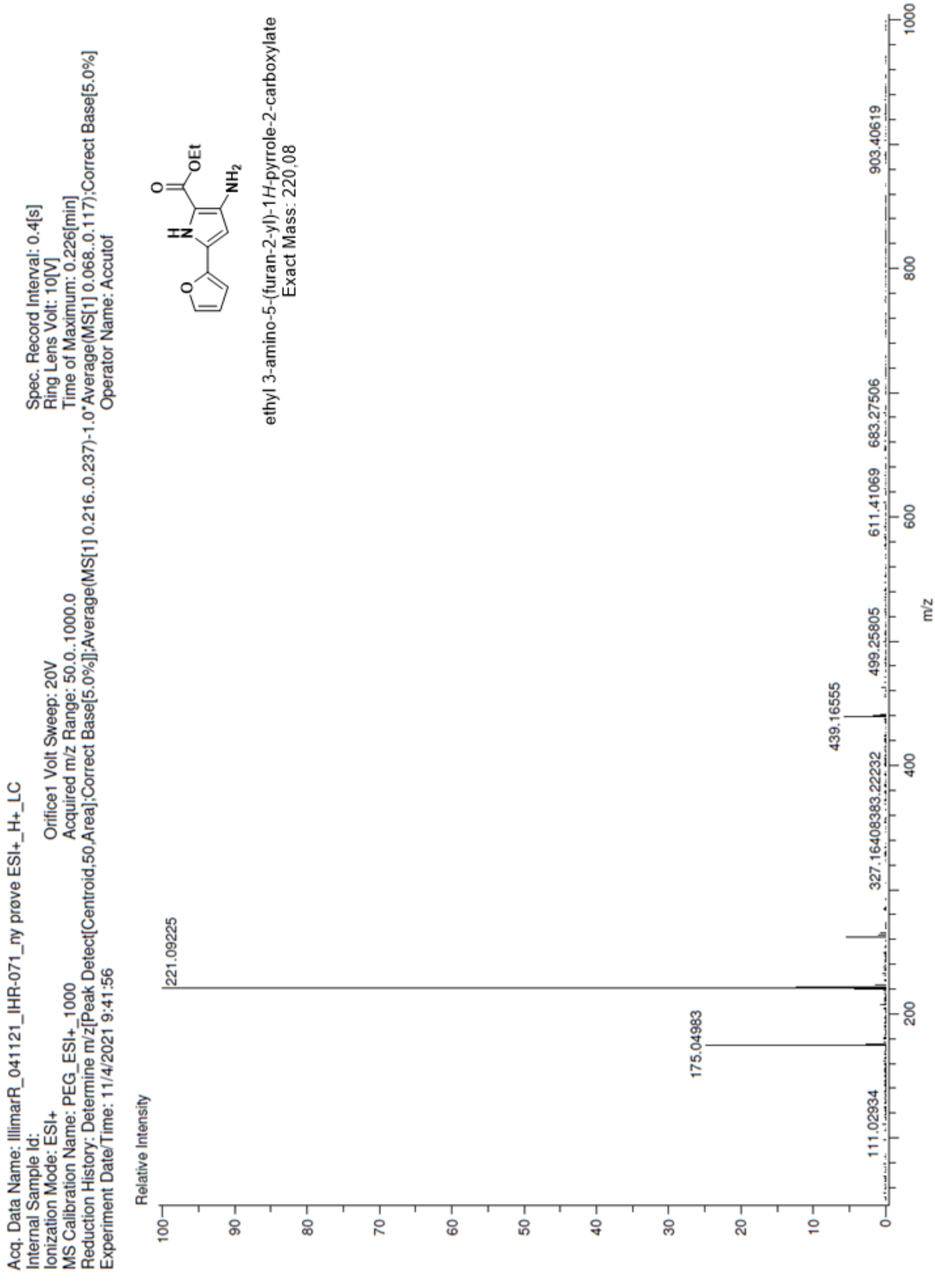
Spectrum 23: ^{1}H NMR of compound ***15b***



Spectrum 24: ^{13}C NMR of compound ***15b***



Spectrum 25: HRMS of compound ***15b***



Acq. Data Name: IllimarR_041121_IHR-071_ny prøve ESI+_H+_LC
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_1000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 0.216..0.237)-1.0*Average(MS[1] 0.068..0.117);Correct Base[5.0%]
Experiment Date/Time: 11/4/2021 9:41:56

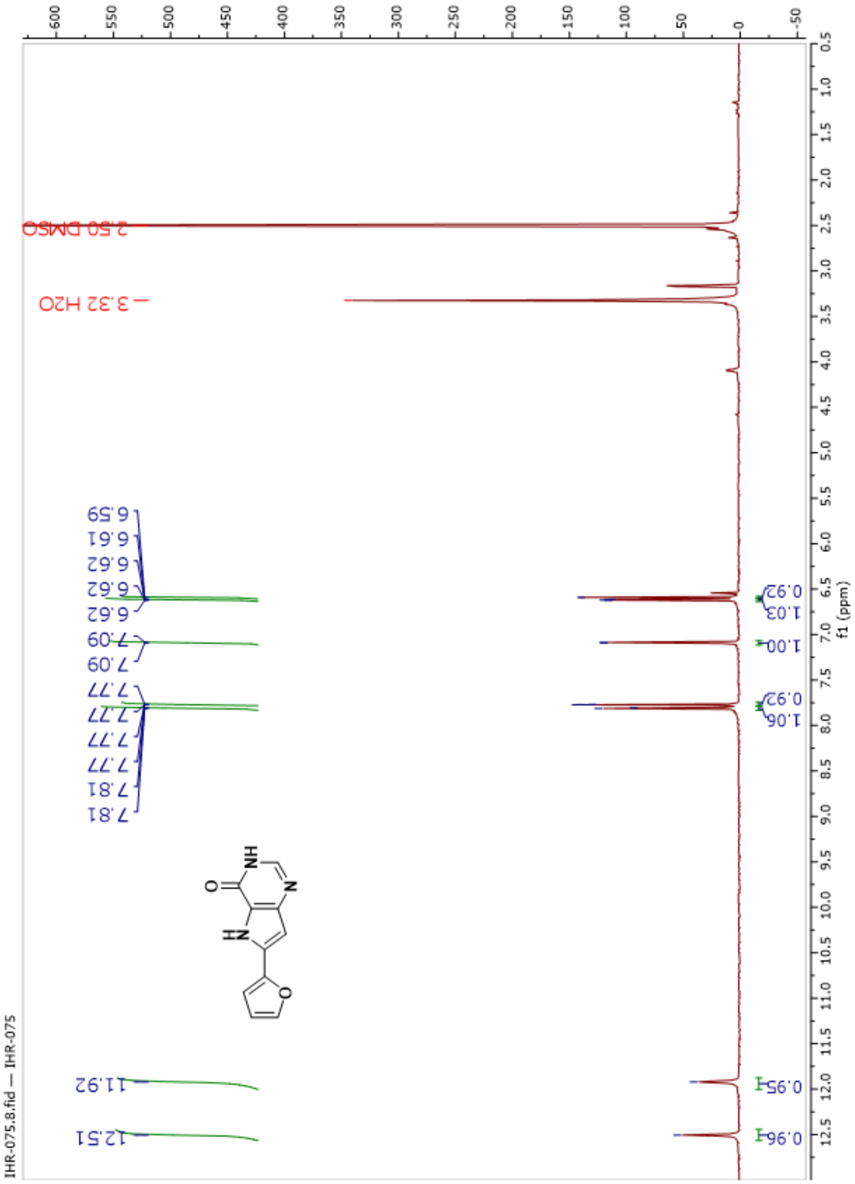
Orifice1 Volt Sweep: 20V
Acquired m/z Range: 50.0..1000.0

Spec. Record Interval: 0.4[s]
Ring Lens Volt: 10[V]
Time of Maximum: 0.226[min]
Operator Name: Accutof

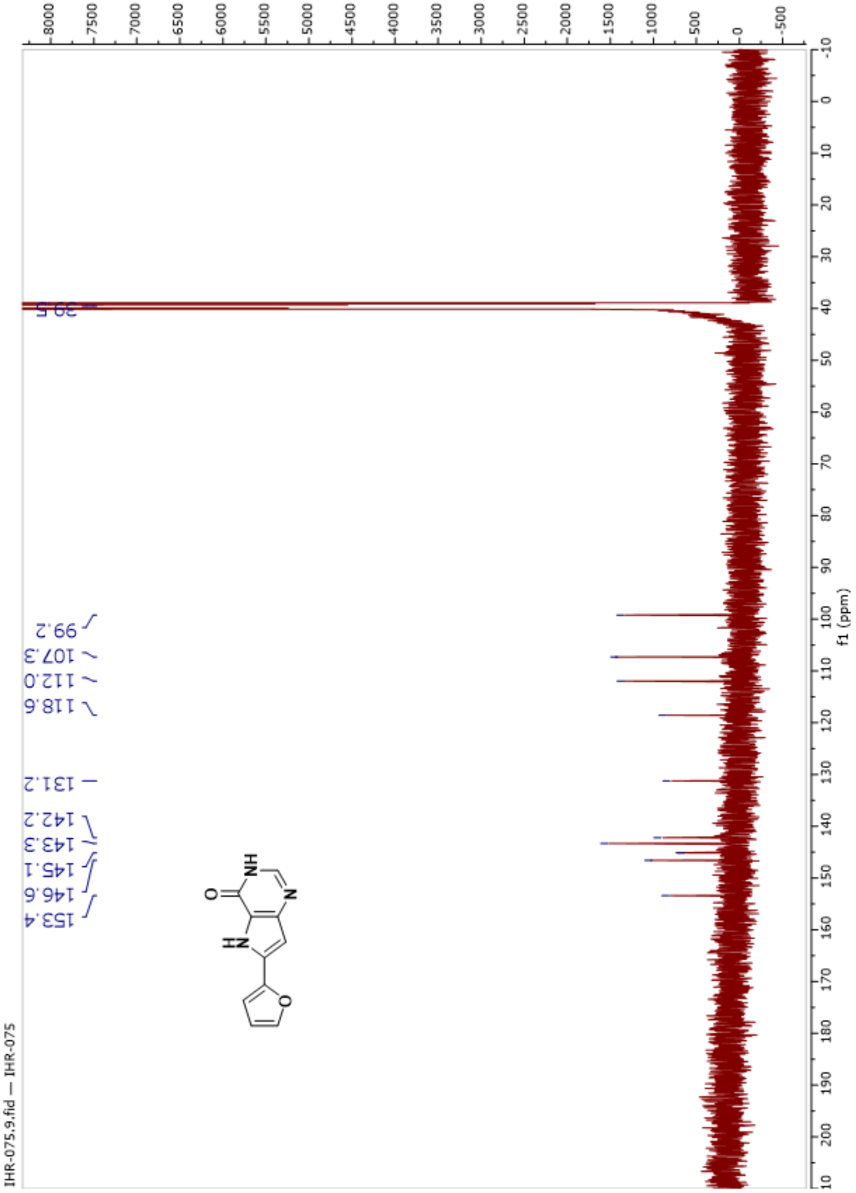
ethyl 3-amino-5-(furan-2-yl)-1*H*-pyrrole-2-carboxylate
Exact Mass: 220,08

Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
221.09225
175.04983
111.02934
327.16408383.22232
439.16555
499.25805
611.41069
683.27506
903.40619
200
400
600
800
1000
m/z
O
H
N
OEt
NH2

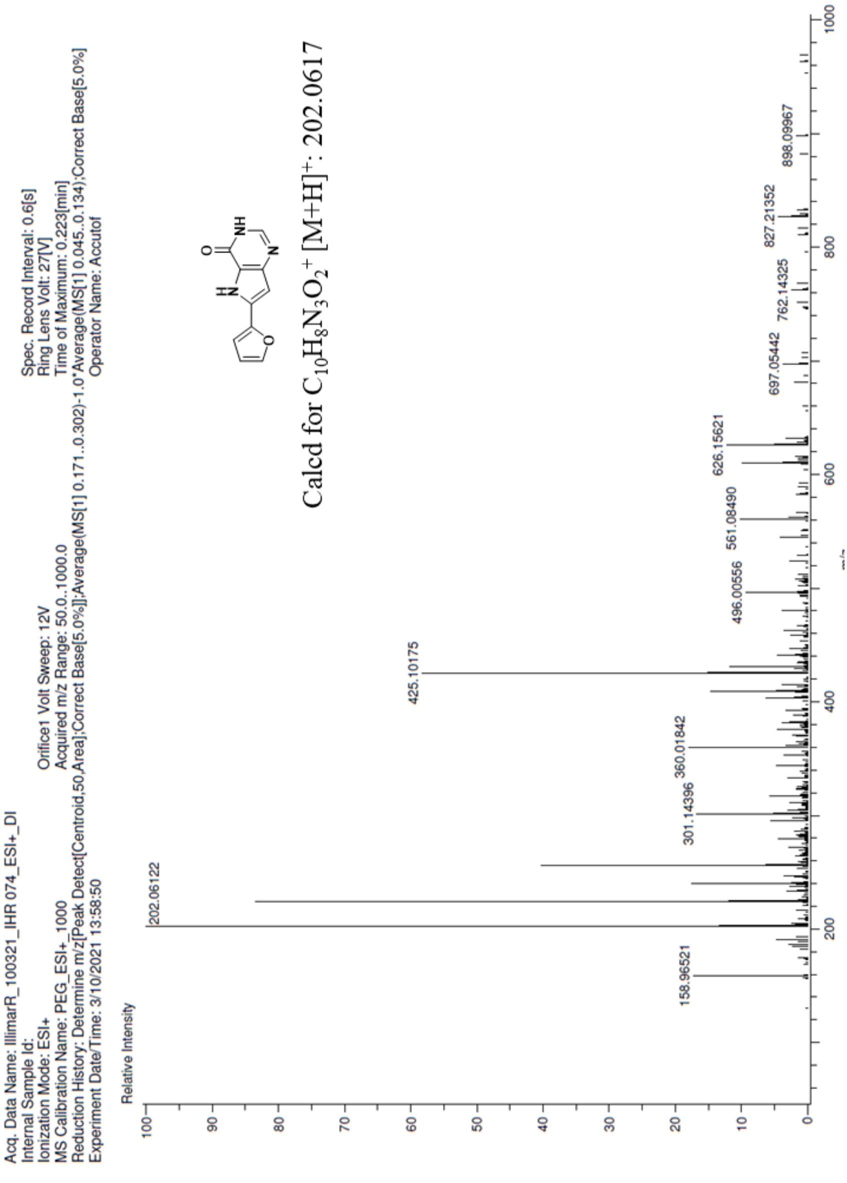
Spectrum 26: ^{1}H NMR of ***RSL-0032***



Spectrum 27: ^{13}C NMR of ***RSL-0032***



Spectrum 28: HRMS of ***RSL-0032***



Acq. Data Name: IllimarR_100321_IHR 074_ESI+_DI
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_1000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 0.171..0.302)-1.0*Average(MS[1] 0.045..0.134);Correct Base[5.0%]
Experiment Date/Time: 3/10/2021 13:58:50

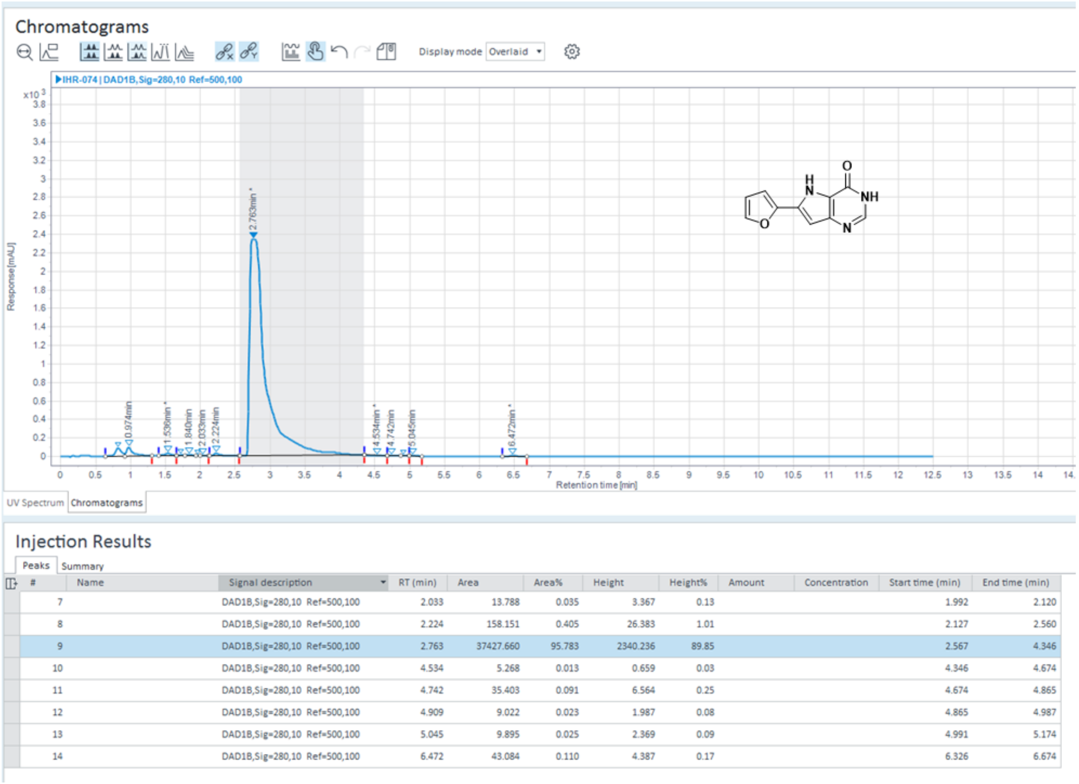
Orifice1 Volt Sweep: 12V
Acquired m/z Range: 50.0..1000.0

Spec. Record Interval: 0.6[s]
Ring Lens Volt: 27[V]
Time of Maximum: 0.223[min]
Operator Name: Accutof

Calcd for $C_{10}H_8N_3O_2^+$ $[M+H]^+$: 202.0617

Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
200
400
600
800
1000
m/z
158.96521
202.06122
301.14396
360.01842
425.10175
496.00556
561.08490
626.15621
697.05442
762.14325
827.21352
898.09967
O
NH
N
H
N
O

Chromatogram 3: HPLC chromatogram of ***RSL-0032***

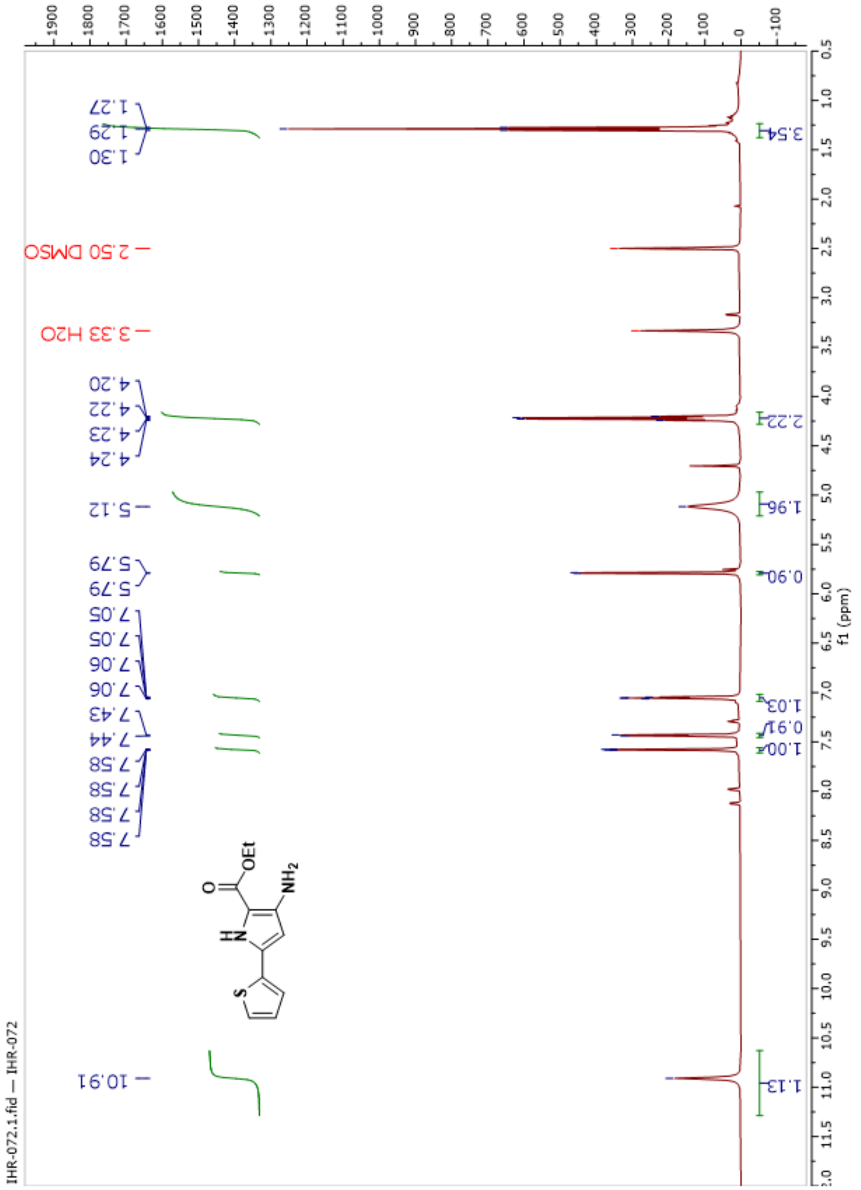

Chromatograms
Display mode Overlaid
▶IHR-074 | DAD1B,Sig=280,10 Ref=500,100
x10^3
3.8
3.6
3.4
3.2
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Response [mAU]
0.974min
1.536min *
1.840min
2.033min
2.224min
2.763min *
4.534min *
4.742min
5.045min
6.472min *
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5
9
9.5
10
10.5
11
11.5
12
12.5
13
13.5
14
14.
Retention time [min]
O
H
N
NH
N
O
UV Spectrum
Chromatograms

Injection Results

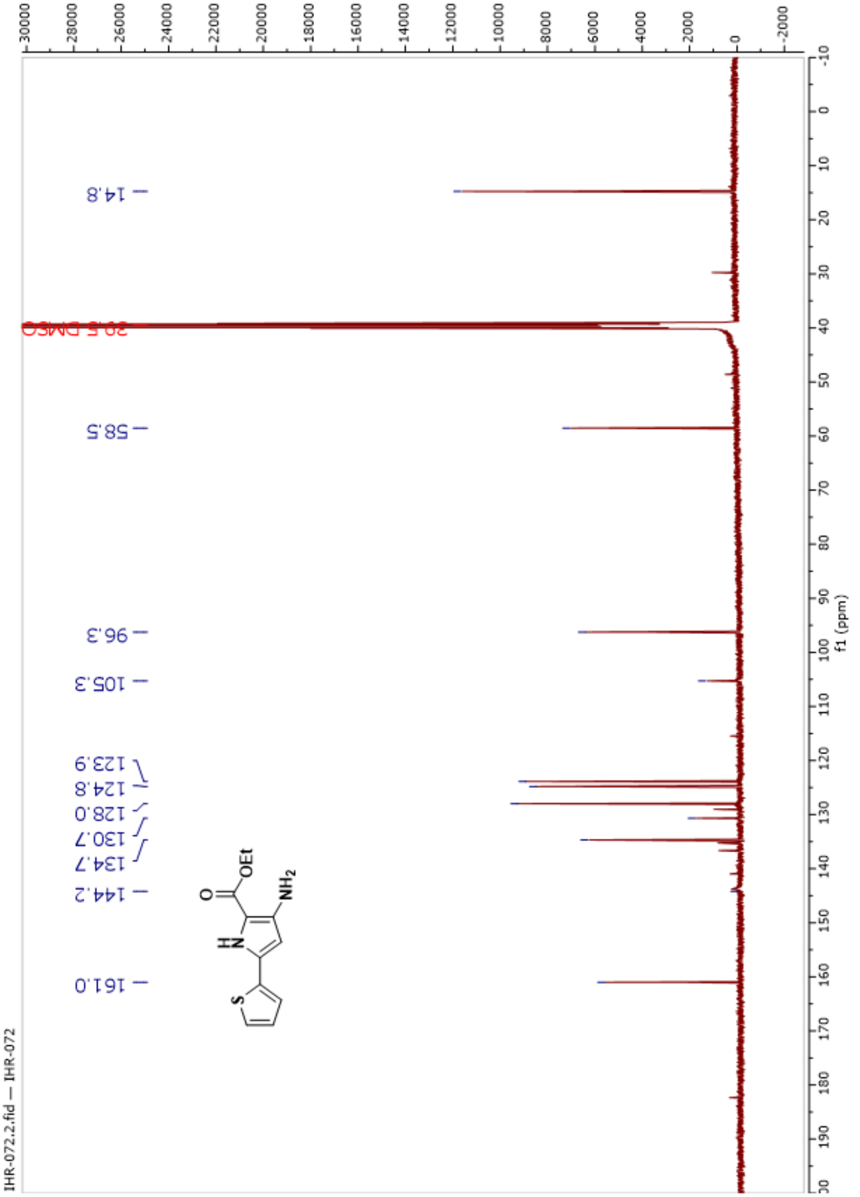
Peaks | Summary

#	Name	Signal description	RT (min)	Area	Area%	Height	Height%	Amount	Concentration	Start time (min)	End time (min)
7		DAD1B,Sig=280,10 Ref=500,100	2.033	13.788	0.035	3.367	0.13			1.992	2.120
8		DAD1B,Sig=280,10 Ref=500,100	2.224	158.151	0.405	26.383	1.01			2.127	2.560
9		DAD1B,Sig=280,10 Ref=500,100	2.763	37427.660	95.783	2340.236	89.85			2.567	4.346
10		DAD1B,Sig=280,10 Ref=500,100	4.534	5.268	0.013	0.659	0.03			4.346	4.674
11		DAD1B,Sig=280,10 Ref=500,100	4.742	35.403	0.091	6.564	0.25			4.674	4.865
12		DAD1B,Sig=280,10 Ref=500,100	4.909	9.022	0.023	1.987	0.08			4.865	4.987
13		DAD1B,Sig=280,10 Ref=500,100	5.045	9.895	0.025	2.369	0.09			4.991	5.174
14		DAD1B,Sig=280,10 Ref=500,100	6.472	43.084	0.110	4.387	0.17			6.326	6.674

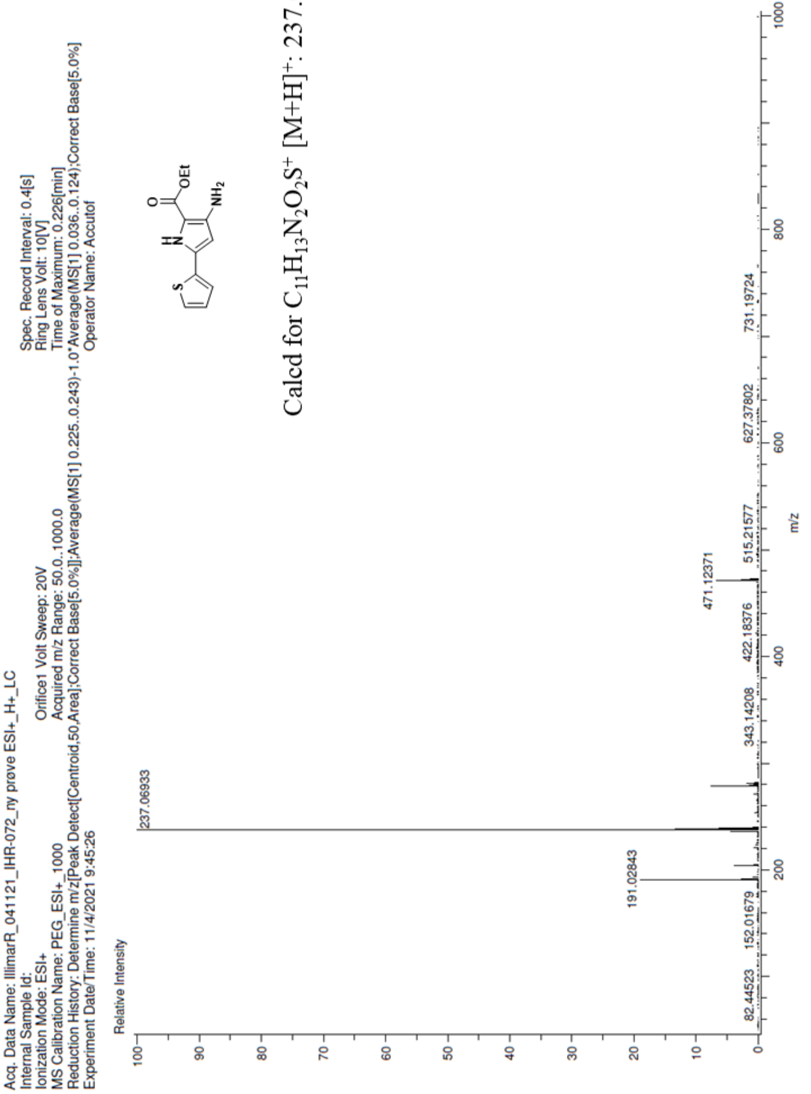
Spectrum 29: ^{1}H NMR of compound ***15c***



Spectrum 30: ^{13}C NMR of compound ***15c***



*Spectrum 31: HRMS of compound **15c***

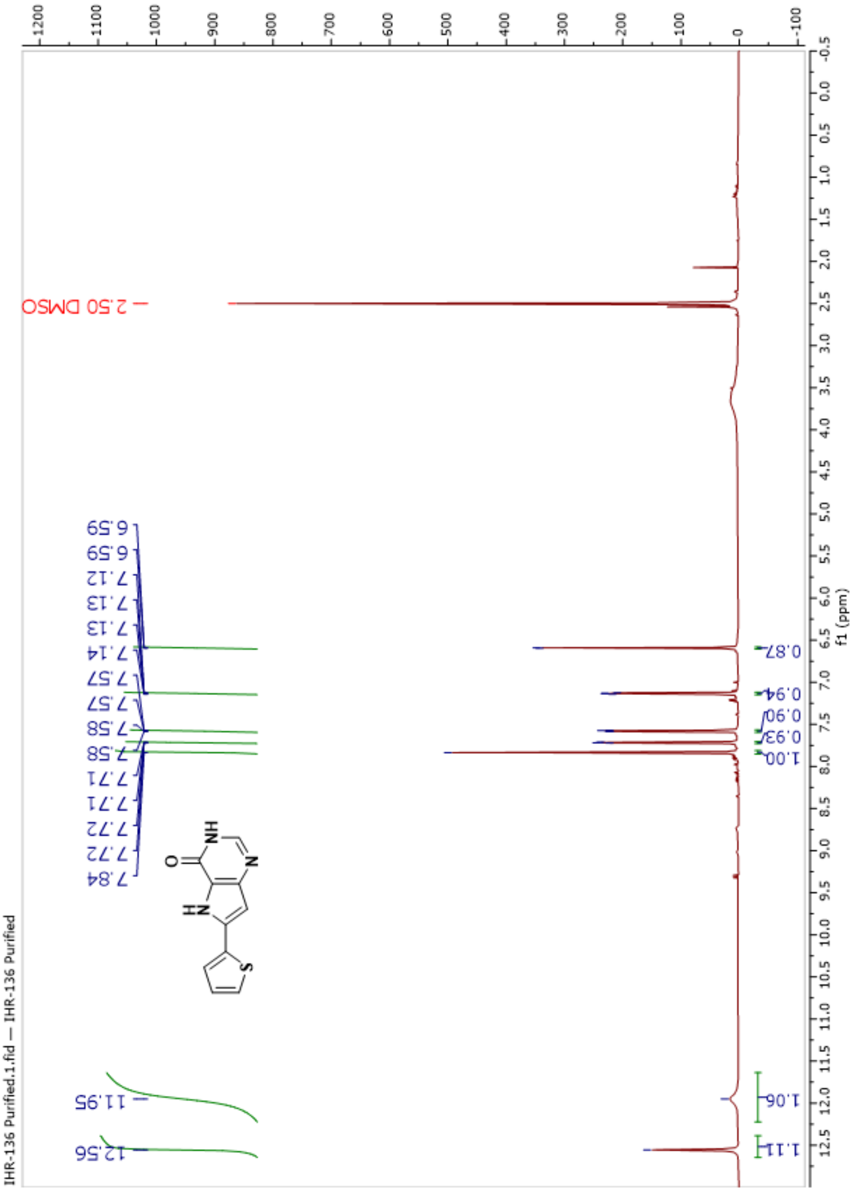


Acq. Data Name: IllimarR_041121_IHR-072_ny prøve ESI+_H+_LC
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_1000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 0.225..0.243)-1.0*Average(MS[1] 0.036..0.124);Correct Base[5.0%]
Experiment Date/Time: 11/4/2021 9:45:26
Orifice1 Volt Sweep: 20V
Acquired m/z Range: 50.0..1000.0
Spec. Record Interval: 0.4[s]
Ring Lens Volt: 10[V]
Time of Maximum: 0.226[min]
Operator Name: Accutof

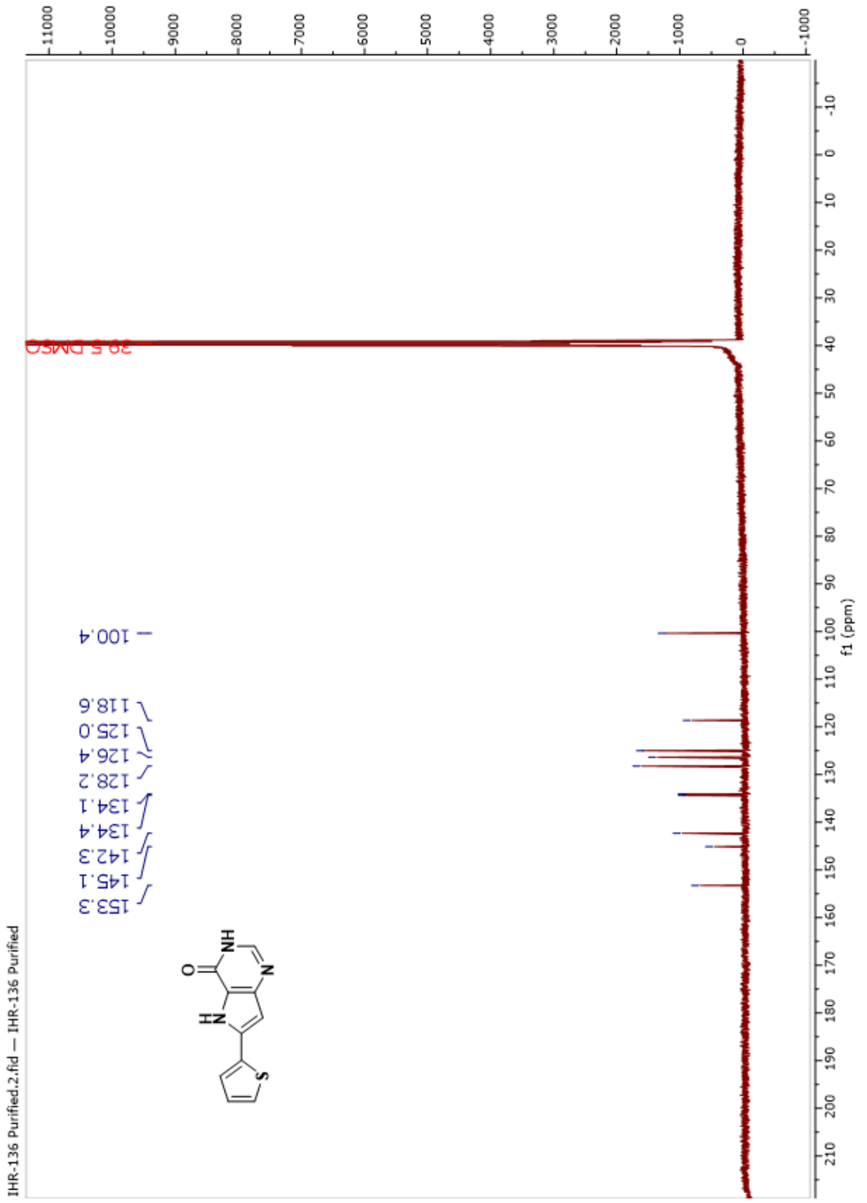
Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
82.44523
152.01679
191.02843
237.06933
343.14208
422.18376
471.12371
515.21577
627.37802
731.19724
200
400
600
800
1000
m/z
S
H
N
O
OEt
NH2

Calcd for $C_{11}H_{13}N_2O_2S^+$ $[M+H]^+$: 237.0698

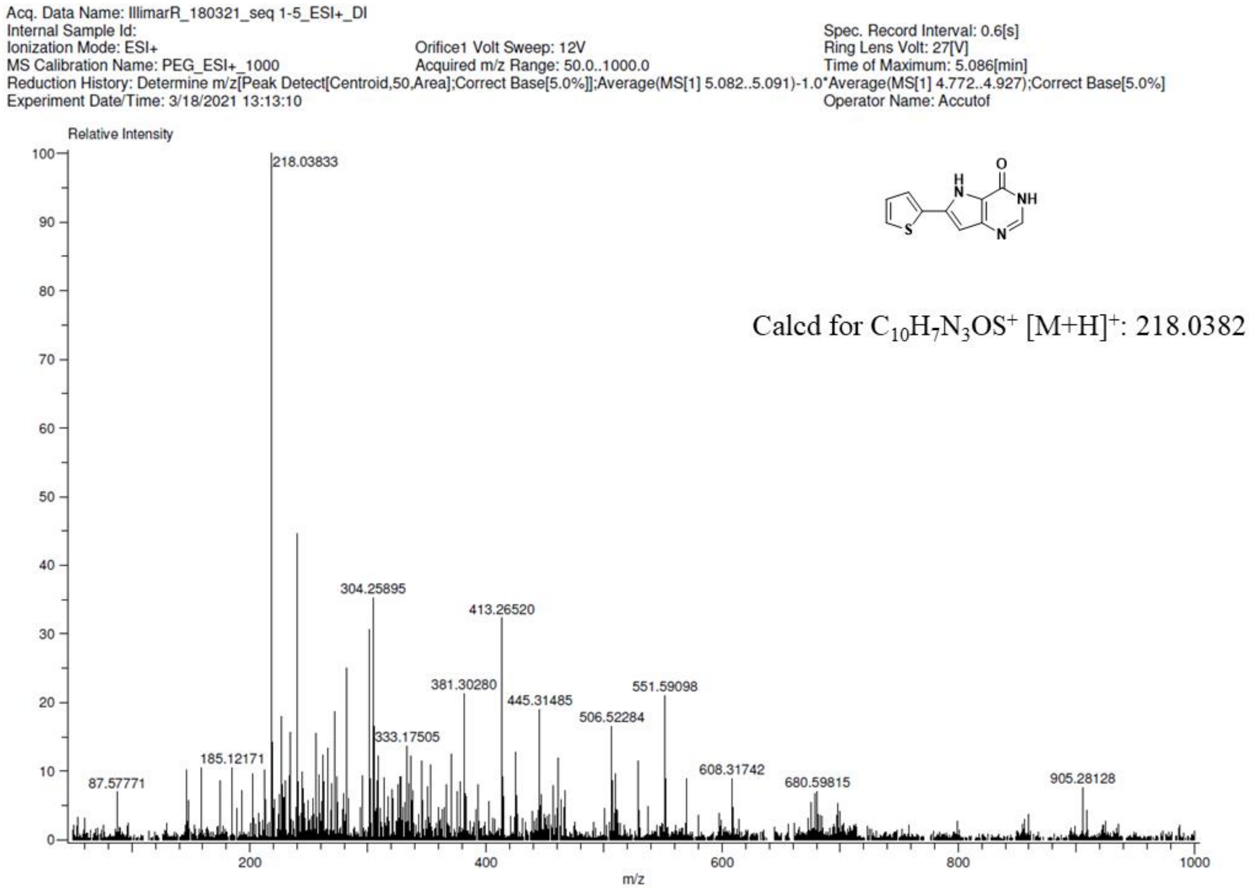
Spectrum 32: ^{1}H NMR of ***RSL-0033***



Spectrum 33: ^{13}C NMR of ***RSL-0033***



Spectrum 34: HRMS of ***RSL-0033***

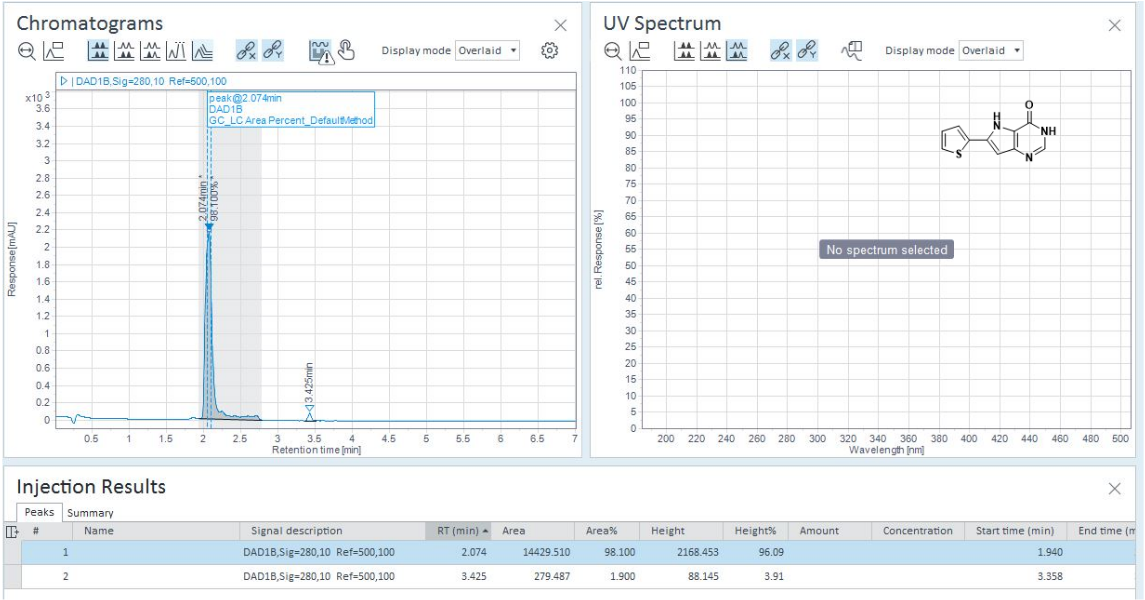


```
Acq. Data Name: IllimarR_180321_seq 1-5_ESI+_DI
Internal Sample Id:                                                                          Spec. Record Interval: 0.6[s]
Ionization Mode: ESI+                           Orifice1 Volt Sweep: 12V                     Ring Lens Volt: 27[V]
MS Calibration Name: PEG_ESI+_1000              Acquired m/z Range: 50.0..1000.0             Time of Maximum: 5.086[min]
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 5.082..5.091)-1.0*Average(MS[1] 4.772..4.927);Correct Base[5.0%]
Experiment Date/Time: 3/18/2021 13:13:10                                                     Operator Name: Accutof
```

Calcd for $C_{10}H_7N_3OS^+$ [M+H]$^+$: 218.0382

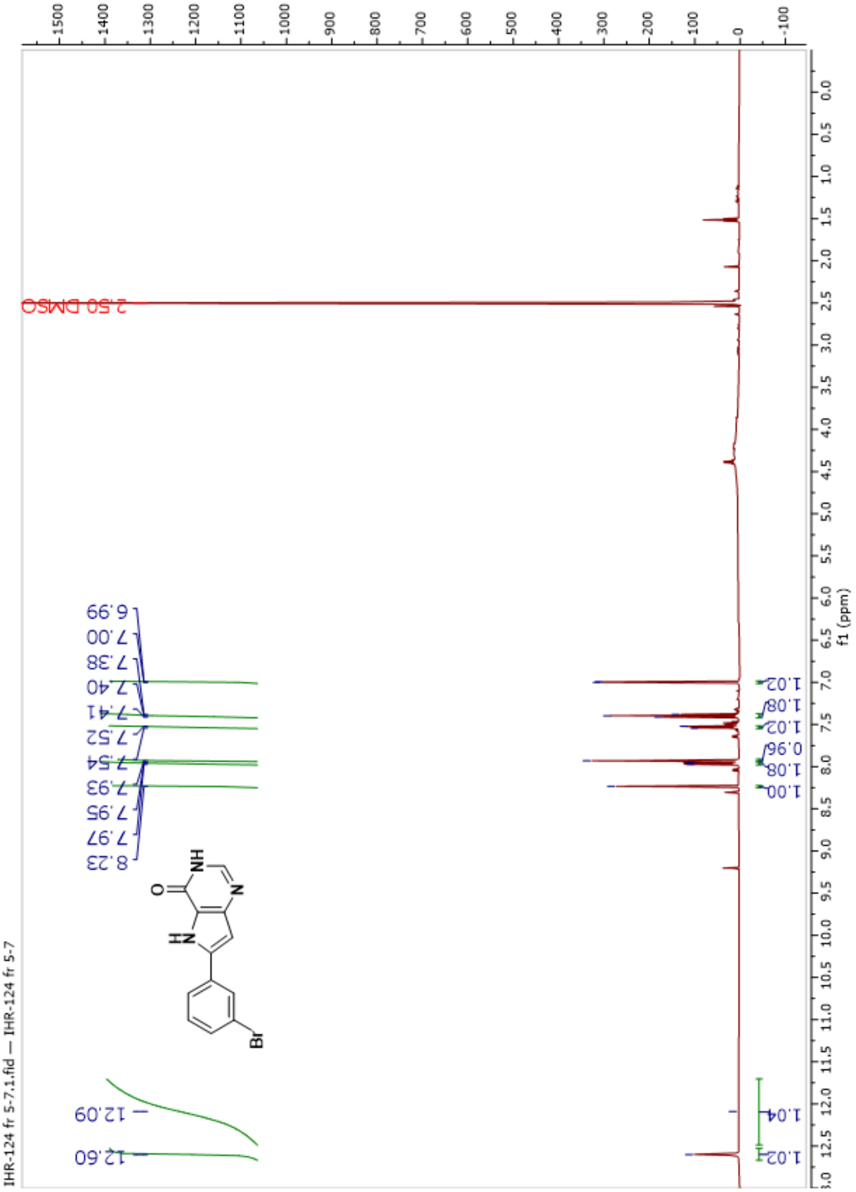
Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
218.03833
304.25895
413.26520
381.30280
445.31485
551.59098
506.52284
333.17505
185.12171
87.57771
608.31742
680.59815
905.28128
200
400
600
800
1000
m/z
H
O
NH
N
S

Chromatogram 4: HPLC of ***RLS-0033***

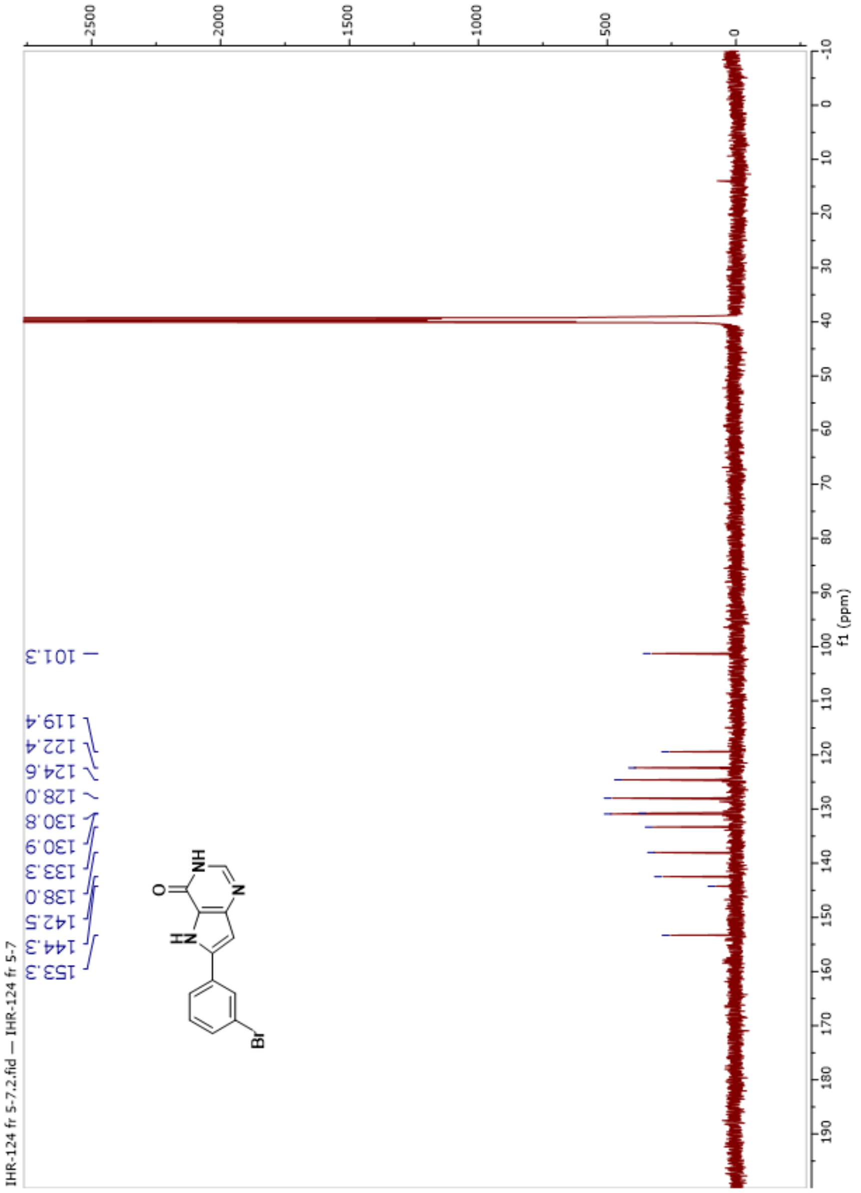

Chromatograms
Display mode
Overlaid
▷ | DAD1B,Sig=280,10 Ref=500,100
peak@2.074min
DAD1B
GC_LC Area Percent_DefaultMethod
2.074min
98.100%
3.425min
Response [mAU]
x10 3
3.6
3.4
3.2
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
Retention time [min]
UV Spectrum
Display mode
Overlaid
rel. Response [%]
110
105
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
200
220
240
260
280
300
320
340
360
380
400
420
440
460
480
500
Wavelength [nm]
No spectrum selected
H
N
O
NH
N
S
Injection Results
Peaks
Summary

#	Name	Signal description	RT (min) ▲	Area	Area%	Height	Height%	Amount	Concentration	Start time (min)	End time (m
1		DAD1B,Sig=280,10 Ref=500,100	2.074	14429.510	98.100	2168.453	96.09			1.940	
2		DAD1B,Sig=280,10 Ref=500,100	3.425	279.487	1.900	88.145	3.91			3.358	

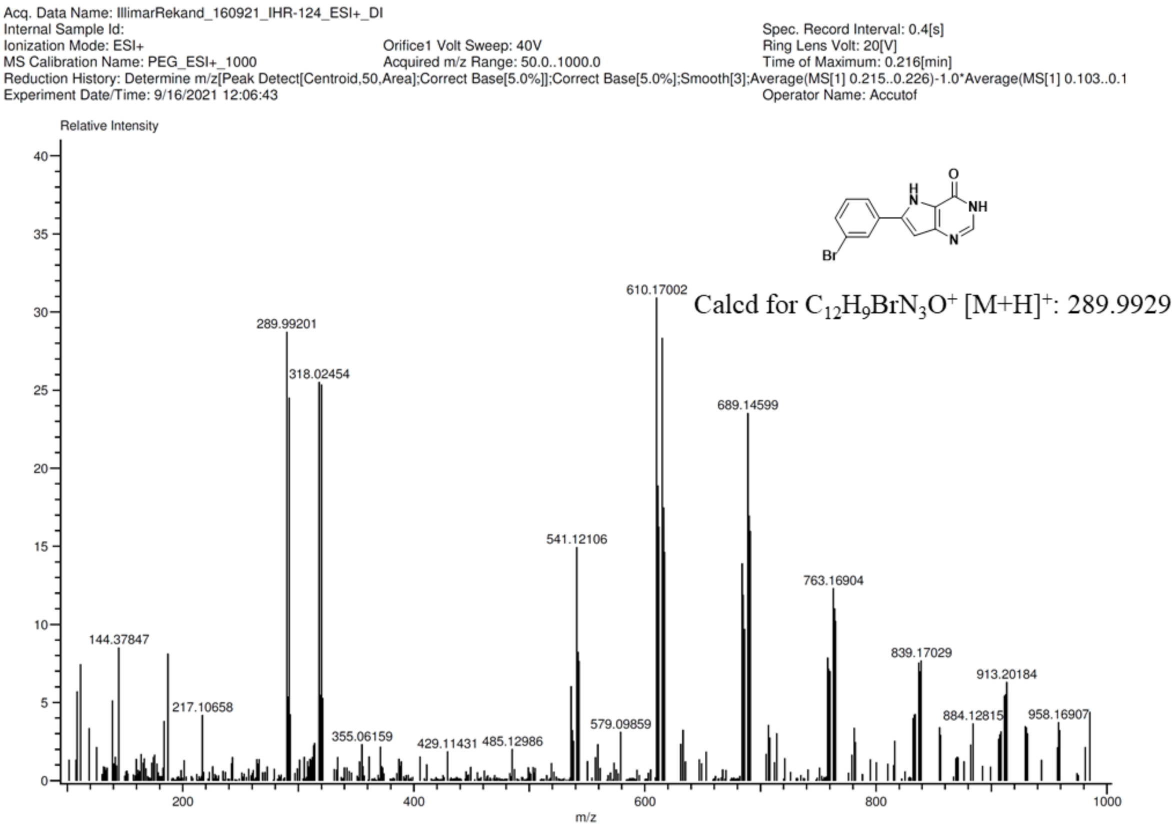
Spectrum 35: ^{1}*H NMR of compound* ***16a***



Spectrum 36: ^{13}C NMR of compound ***16a***



Spectrum 37: HRMS of compound ***16a***

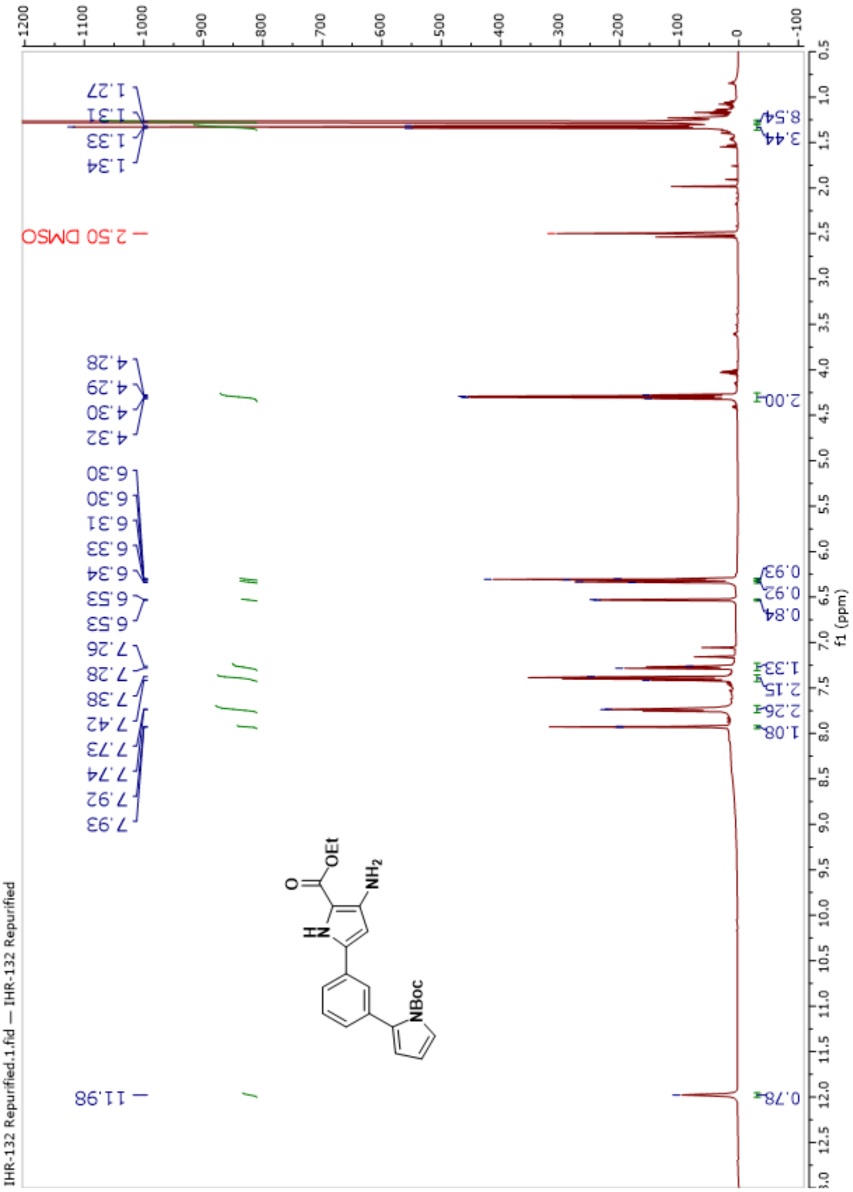


Acq. Data Name: IllimarRekand_160921_IHR-124_ESI+_DI
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_1000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Correct Base[5.0%];Smooth[3];Average(MS[1] 0.215..0.226)-1.0*Average(MS[1] 0.103..0.1
Experiment Date/Time: 9/16/2021 12:06:43
Orifice1 Volt Sweep: 40V
Acquired m/z Range: 50.0..1000.0
Spec. Record Interval: 0.4[s]
Ring Lens Volt: 20[V]
Time of Maximum: 0.216[min]
Operator Name: Accutof

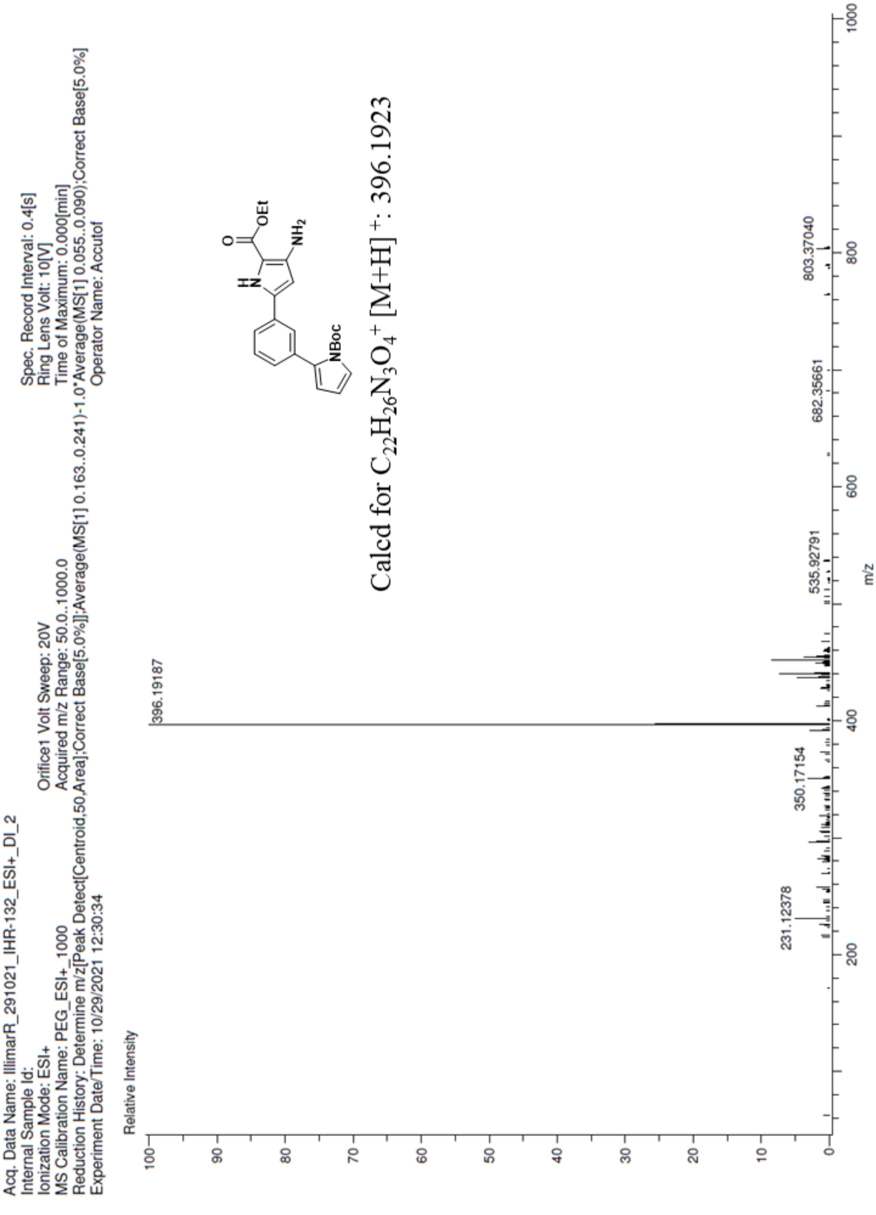
Calcd for $C_{12}H_9BrN_3O^+$ $[M+H]^+$: 289.9929

Relative Intensity
40
35
30
25
20
15
10
5
0
200
400
600
800
1000
m/z
144.37847
217.10658
289.99201
318.02454
355.06159
429.11431
485.12986
541.12106
579.09859
610.17002
689.14599
763.16904
839.17029
884.12815
913.20184
958.16907
H
N
O
NH
N
Br

*Spectrum 38: ^{1}H NMR of compound **16b***



Spectrum 39: HRMS of compound ***16b***

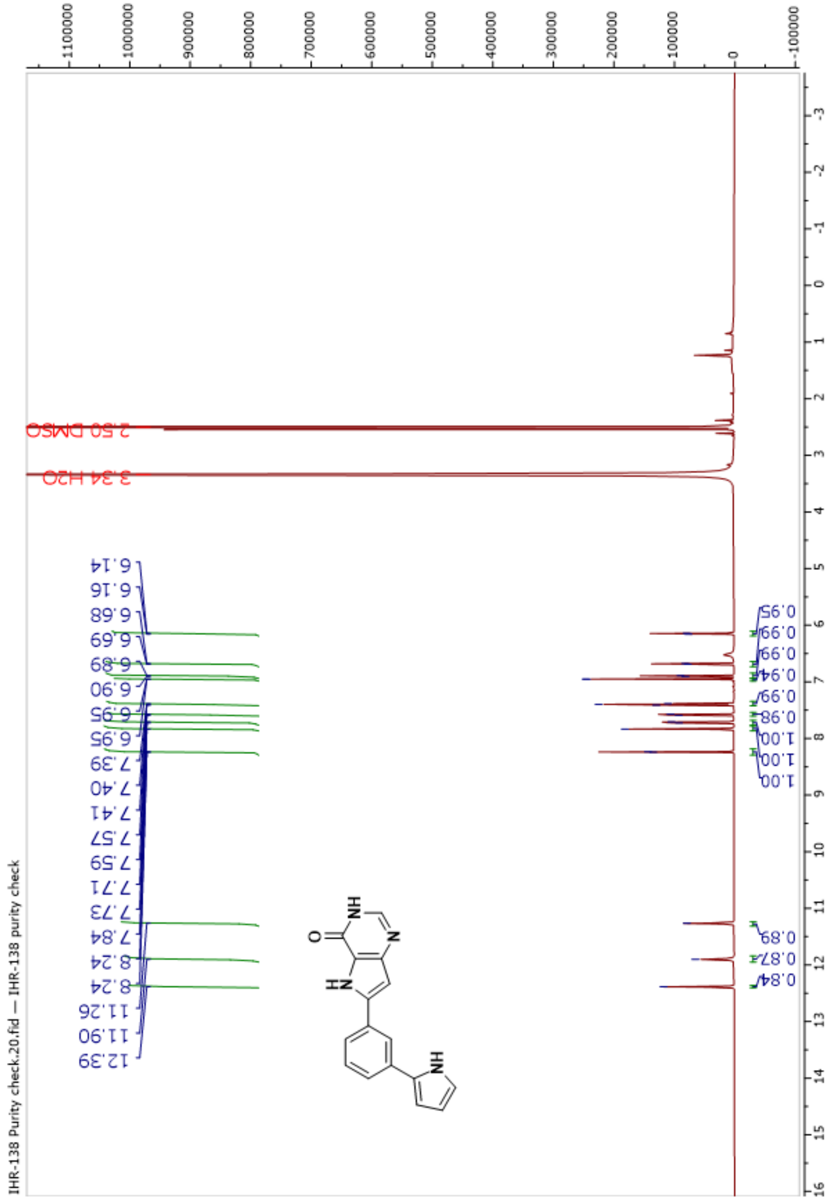


Acq. Data Name: IllimarR_291021_IHR-132_ESI+_DI_2
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_1000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 0.163..0.241)-1.0*Average(MS[1] 0.055..0.090);Correct Base[5.0%]
Experiment Date/Time: 10/29/2021 12:30:34
Orifice1 Volt Sweep: 20V
Acquired m/z Range: 50.0..1000.0
Spec. Record Interval: 0.4[s]
Ring Lens Volt: 10[V]
Time of Maximum: 0.000[min]
Operator Name: Accutof

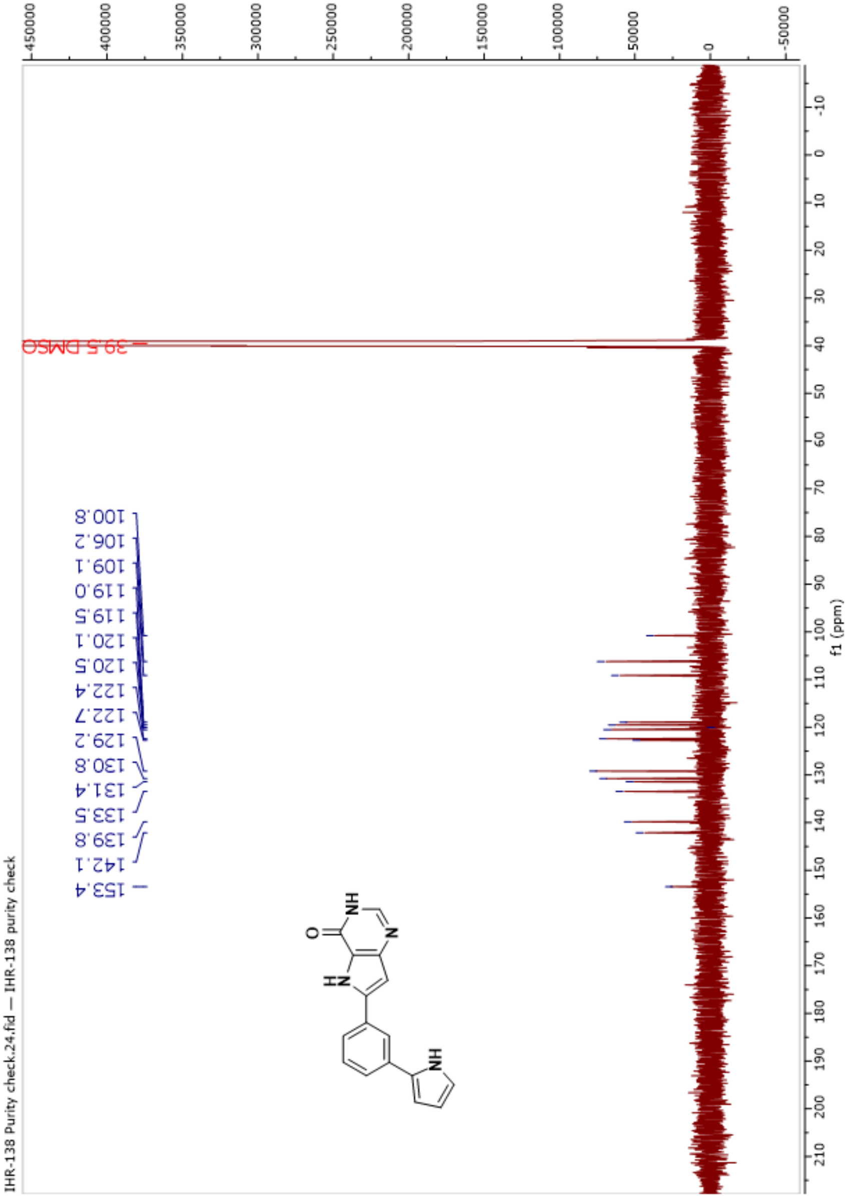
Calcd for $C_{22}H_{26}N_3O_4^+$ $[M+H]^+$: 396.1923

Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
m/z
200
400
600
800
1000
231.12378
350.17154
396.19187
535.92791
682.35661
803.37040
O
OEt
NH2
H
N
NBoc

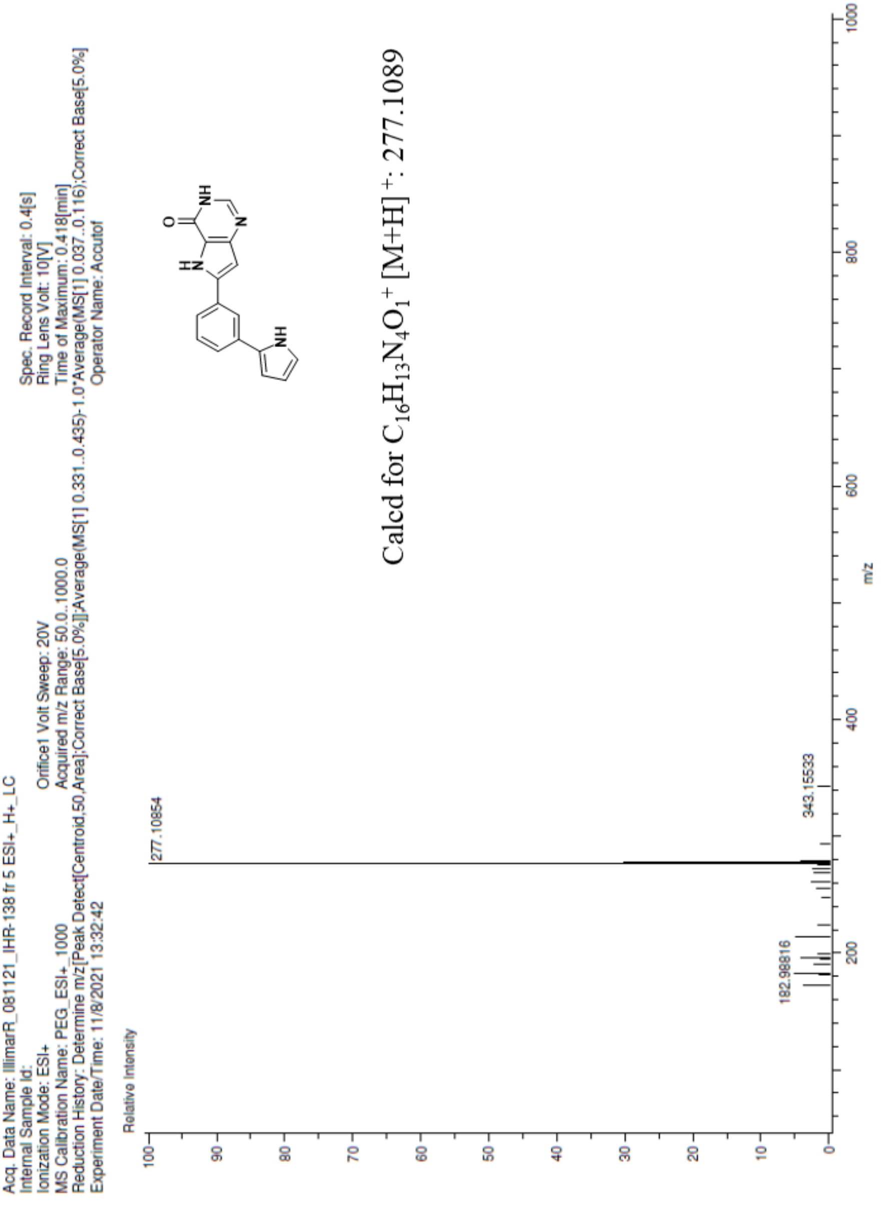
Spectrum 40: ^{1}H NMR of **RSL-0034**



Spectrum 41: ^{13}C NMR of ***RSL-0034***



Spectrum 42: HRMS of ***RLS-0034***



Acq. Data Name: IllimarR_081121_IHR-138 fr 5 ESI+_H+_LC
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_1000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 0.331..0.435)-1.0*Average(MS[1] 0.037..0.116);Correct Base[5.0%]
Experiment Date/Time: 11/8/2021 13:32:42

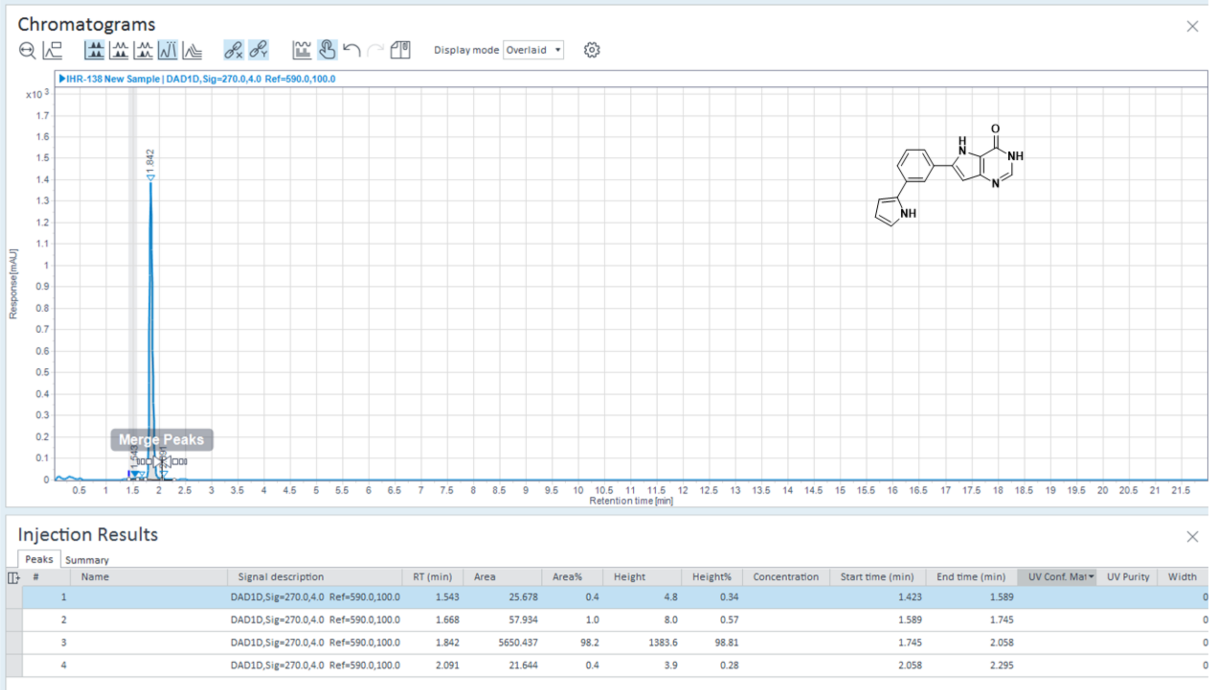
Orifice1 Volt Sweep: 20V
Acquired m/z Range: 50.0..1000.0

Spec. Record Interval: 0.4[s]
Ring Lens Volt: 10[V]
Time of Maximum: 0.418[min]
Operator Name: Accutof

Calcd for $C_{16}H_{13}N_4O_1^+$ $[M+H]^+$: 277.1089

Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
277.10854
343.15533
182.98816
200
400
600
800
1000
m/z
O
NH
N
H
N
NH

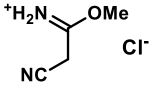
Chromatogram 5: HPLC of ***RSL-0034***


Chromatograms
Display mode Overlaid
▶IHR-138 New Sample | DAD1D,Sig=270.0,4.0 Ref=590.0,100.0
x10 3
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Response [mAU]
1.842
1.543
2.091
Merge Peaks
0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 10 10.5 11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 16 16.5 17 17.5 18 18.5 19 19.5 20 20.5 21 21.5
Retention time [min]
H
N
O
NH
N
NH
Injection Results
Peaks
Summary

#	Name	Signal description	RT (min)	Area	Area%	Height	Height%	Concentration	Start time (min)	End time (min)	UV Conf. Mat	UV Purity	Width
1		DAD1D,Sig=270.0,4.0 Ref=590.0,100.0	1.543	25.678	0.4	4.8	0.34		1.423	1.589			0
2		DAD1D,Sig=270.0,4.0 Ref=590.0,100.0	1.668	57.934	1.0	8.0	0.57		1.589	1.745			0
3		DAD1D,Sig=270.0,4.0 Ref=590.0,100.0	1.842	5650.437	98.2	1383.6	98.81		1.745	2.058			0
4		DAD1D,Sig=270.0,4.0 Ref=590.0,100.0	2.091	21.644	0.4	3.9	0.28		2.058	2.295			0

6.3 Experiments leading up to and including **RSL-0036**

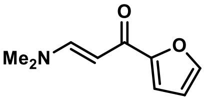
2-Cyanoethanimidic acid methyl ester hydrochloride (**21)**



Malononitrile (10.5 g, 159 mmol) was dissolved in a mixture of methanol (15 mL) and diethyl ether (500 mL) and the mixture was cooled to 0 °C. A gentle stream of HCl gas (generated by dripping concentrated sulfuric acid onto NaCl) was bubbled through the mixture over 2.5 hours. Afterwards, dichloromethane (500 mL) was added, and the precipitate was filtered off, washed with a 1:1 (v/v) mixture of dichloromethane/diethyl ether and dried under vacuum to give the title compound was isolated as a beige solid (7.38 g, 35 % yield). Melting point: 103-107 °C.

The analytical data matched those previously published.[150]

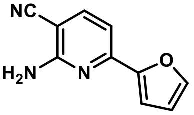
3-(Dimethylamino)-1-(furan-2-yl)prop-2-en-1-one (**22**)



A mixture of 2-acetyl furane (10.4 g, 94.8 mmol) and *N,N*-dimethylformamide dimethyl acetal (13.7 g, 114 mmol) was heated to reflux and stirred for 5.5 hours. After cooling, the mixture was concentrated *in vacuo* and the residue was dissolved in diethyl ether, and *n*-pentane was added to give a precipitate. The precipitate was isolated by filtration on a Büchner funnel and washed with a 1/1-mixture (*v/v*) of diethyl ether and *n*-pentane, to give the title compound as a brown crystalline solid (13.7 g, 87%).

^{1}H-NMR (500 MHz, $(CD_3)_2SO$): δ = 7.78 (dd, *J* = 1.7, 0.8, 1H), 7.66 (d, *J* = 12.5, 1H), 7.10 (dd, *J* = 3.4, 0.8, 1H), 6.59 (dd, *J* = 1.7, 3.5, 1H), 5.63 (d, *J* = 12.4, 1H), 3.13 (s, 3H), 2.88 (s, 3H); ^{13}C-NMR (125 MHz, $(CD_3)_2SO$): δ = 175.5, 154.5, 153.3, 144.8, 113.1, 111.9, 90.6, 44.5, 37.1
The analytical data matched those previously published.[150]

2-Amino-6-(furan-2-yl)nicotinonitrile (**23**)

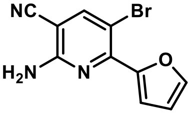


2-cyanoimidic methyl ester **21** (2.02 g, 12.3 mmol) and ammonium acetate (5.04 g, 23.7 mmol) were dissolved in methanol (50 mL) and the resulting mixture was heated to reflux for 15 minutes. Enone **16 w**as added and heating was continued overnight. The solvent was removed under reduced pressure and cold water (50 mL) was added to the residue. The precipitate was isolated by filtration on a Büchner funnel to give the title compound as a beige solid (1.45 g, 63%).

^{1}H NMR (500 MHz, $(CD_3)_2SO$): δ = 7.92 (d, *J* = 8.1, 1H), 7.88 (dd, *J* = 0.8, 1.8, 1H), 7.13 (dd, *J* = 3.5, 0.8, 1H), 7.01 (d, *J* = 8.0, 1H), 6.94 (br. s, 2H), 6.67 (dd, *J* = 3.5, 1.8, 1H)**;** ^{13}C-NMR (125 MHz, $(CD_3)_2SO$): δ = 159.7, 152.2, 150.7, 145.3, 143.1, 117.2, 112.6, 111.4, 106.7, 87.5

The analytical data matched those previously published.[150]

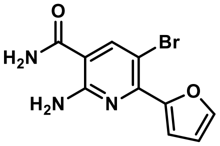
2-Amino-5-bromo-6-(furan-2-yl)nicotinonitrile (**24**)



Pyridine **23** (1.45 g, 7.81 mmol) and N-bromosuccinimide (1.41 g, 7.90 mmol) were dissolved in DMF (30 mL) and the resulting mixture was stirred overnight. The mixture was diluted with ethyl acetate (200 mL) and washed twice with saturated $NaHCO_3$ (2 x 250 mL) and saturated NaCl (250 mL). The combined aqueous phases were extracted with ethyl acetate (4 x 100 mL), and the combined organic phases were concentrated in vacuo to give a grey solid. The crude material was purified using a Biotage Sfär silica gel cartridge (20 µm, 50 g) eluting with EtOAc in hexane (20% to 40% over 30 minutes and 30 mL/min flow rate). The peak that eluted at approx. 32% was collected, and after removal of solvent the title compound was isolated as a beige solid (1.34 g, 64%)

^{1}H NMR (500 MHz, $(CD_3)_2SO$) δ = 8.26 (s, 1H), 7.95 (dd, *J* = 1.7, 0.7, 1H), 7.43 (dd, *J* = 3.5, 0.7, 1H), 7.17 (s, 2H), 6.72 (dd, *J* = 3.5, 1.7, 1H); ^{13}C-NMR ($(CD_3)_2SO$, 125 MHz): δ = 157.8, 150.0, 148.1, 147.1, 145.3, 115.6, 115.5, 112.2, 99.6, 89.7
The analytical data matched those previously published.[150]

2-Amino-5-bromo-6-(furan-2-yl)nicotinamide (**19)**

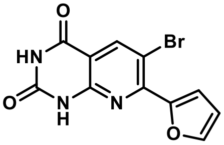


Nitrile **23** (432 mg, 1.63 mmol) was dissolved in ethanol (8 mL) with ethanol (8 mL) in a microwave tube and saturated aqueous ammonia (8 mL) and hydrogen peroxide (449 µl, 50% *v/v*) were added. The tube was sealed and heated to 50 °C for 2.5 h. After cooling to room temperature, the reaction mixture was diluted with ethyl acetate (5 mL) and extracted with ethyl acetate (3 x 50 ml). The combined organic phases were dried over Na_2SO_4, and the solvent was removed *in vacuo* to give the title compound as a yellow solid (437 mg, 95%)

^{1}H NMR (500 MHz, $(CD_3)_2SO$): δ = 8.27 (s, 1H), 8.08 (bs, 1H), 7.90 (dd, *J* = 1.7, 0.7), 7.45 (bs, 3H), 7.34 (dd, 3.5, 0.7, 1H), 6.69 (dd, *J* = 3.5, 1.8); ^{13}C-NMR ($(CD_3)_2SO$, 125 MHz): δ = 168.2, 157.1, 150.6, 146.8, 144.5, 142.4, 114.1, 111.9, 109.0, 99.6

The analytical data matched those previously published.[150]

6-Bromo-7-(furan-2-yl)pyrido[2,3-d]pyrimidine-2,4(1H,3H)-dione (**18)**



The title compound was prepared using one of two methods:

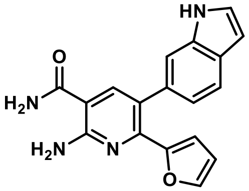
Method 1: Aminonitrile **19** (101 mg, 0.36 mmol) was dissolved in DMSO (2 mL), and NaH, 60% dispersion in oil, (40.3 mg, 1.01 mmol) was added. The mixture was stirred for 30 minutes at room temperature, before carbonyl diimidazole (63.9 mg, 0.39 mmol), dissolved in DMF (1 mL), was added to the reaction mixture and stirred for 2 h at 90 °C. After cooling to room temperature, 1 M HCl (100 mL) was added to the reaction mixture. The reaction mixture was extracted with EtOAc (4 x 25 mL) and H_2O (25 mL), and the combined organic layers were dried over Na_2SO_4. The crude material was then further purified using a Biotage Sfär C18 cartridge (20 µm, 12 g) eluting with acetonitrile in water, both containing 0.1% TFA (10 to 60% over 20 min, 15 mL/min). The product was eluted at 37% acetonitrile and the title compound was isolated as an orange solid (43 mg, 38%) after lyophilization.

Method 2: Aminonitrile **19** (59.6 mg, 0.212 mmol) and carbonyl diimidazole (42.5 mg, 0.262 mmol) were dissolved in DMF (2 mL) and 1,8-diazabicyclo(5.4.0)undec-7-ene (3.4 mg, 0.0223 mmol) dissolved in DMF (340 µL) was added. The resulting mixture was heated to 110 °C for 8 h. The reaction mixture was extracted (4 x 25 mL ethyl acetate) and the combined organic phases were dried over Na_2SO_4. The crude material was then further purified using a Biotage Sfär C18 cartridge (20 µm, 12 g) eluting with acetonitrile in water, both containing 0.1% TFA (10 to 60% over 20 min, 15 mL/min). The product was eluted at 38% acetonitrile and the title compound was isolated as an orange solid (27 mg, 49%) after lyophilization.

^{1}H NMR (500 MHz, $(CD_3)_2SO$**)** δ = 11.87, (s, 1H), 11.55 (s, 1H), 8.40 (d, *J* = 0.5, 1H), 8.04, (dd, *J* = 0.7, 1.7), 7.58 (dd, *J* = 0.7, 3.6), 6.78 (dd, *J* = 1.7, 3.6) ^{13}C-NMR ($(CD_3)_2SO$, 125 MHz) δ = 161.5, 151.3, 150.3, 149.8, 146.5, 142.00, 117.0, 113.0, 110.3, 108.4

The analytical data matched those previously published.[150]

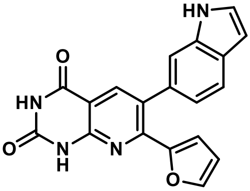
7-(Furan-2-yl)-6-(1H-indol-6-yl)pyrido[2,3-d]pyrimidine-2,4(1H,3H)-dione (**25**)



Bromide **19** (97.6 mmol, 0.347 mmol), 6-indolyl boronic acid (103.6 mg, 0.644 mmol) and Cs_2CO_3(280 mg, 0.860 mmol) were dissolved in 1,4-dioxane (4 mL). The solution was purged and degassed under high vacuum, repeated three times. $Pd(dppf)Cl_2 \cdot CH_2Cl_2$ (37.2 mg, 0.0720 mmol) was subsequently added, and the reaction mixture was heated to 90 °C for 2 h, after which another portion of 6-indolyl boronic acid (57 mg, 0.35 mmol) was added. The reaction mixture was allowed to stir for 2 h at 90 °C, before it was cooled to room temperature and separated between water (35 mL) and ethyl acetate (3 x 35 mL). The combined organic layers were dried using Na_2SO_4, and the crude material was then further purified using a Biotage Sfär C18 cartridge (20 μm, 12 g) eluting with acetonitrile in water, both containing 0.1% TFA (10 to 50% over 20 min, 15 mL/min). The product was eluted at 25% acetonitrile and the title compound was isolated as an orange solid (94 mg, 85%) after lyophilization.).

^{1}H NMR (500 MHz, DMSO) δ = 11.22 (s, 1H), 8.21 (br. s, 2H), 7.81 (s, 1H), 7.59 (d, *J* = 8.0, 1H), 7.41 (t, *J* = 2.8), 7.35 (s, 1H), 6.92 (dd, *J* = 1.4, 8.0), 6.49 (m, 1H), 6.46 (dd, *J* = 1.4, 3.2), 5.92, (dd, *J* = 0.7, 3.4 ^{13}C-NMR ($(CD_3)_2SO$, 125 MHz) 13C-NMR data of sufficient quality was not obtained for this structure; HRMS (ESI) m/z calcd for $C_{18}H_{15}N_4O_2^+$ [M+H]$^+$: 319.1195, found: 319.1191

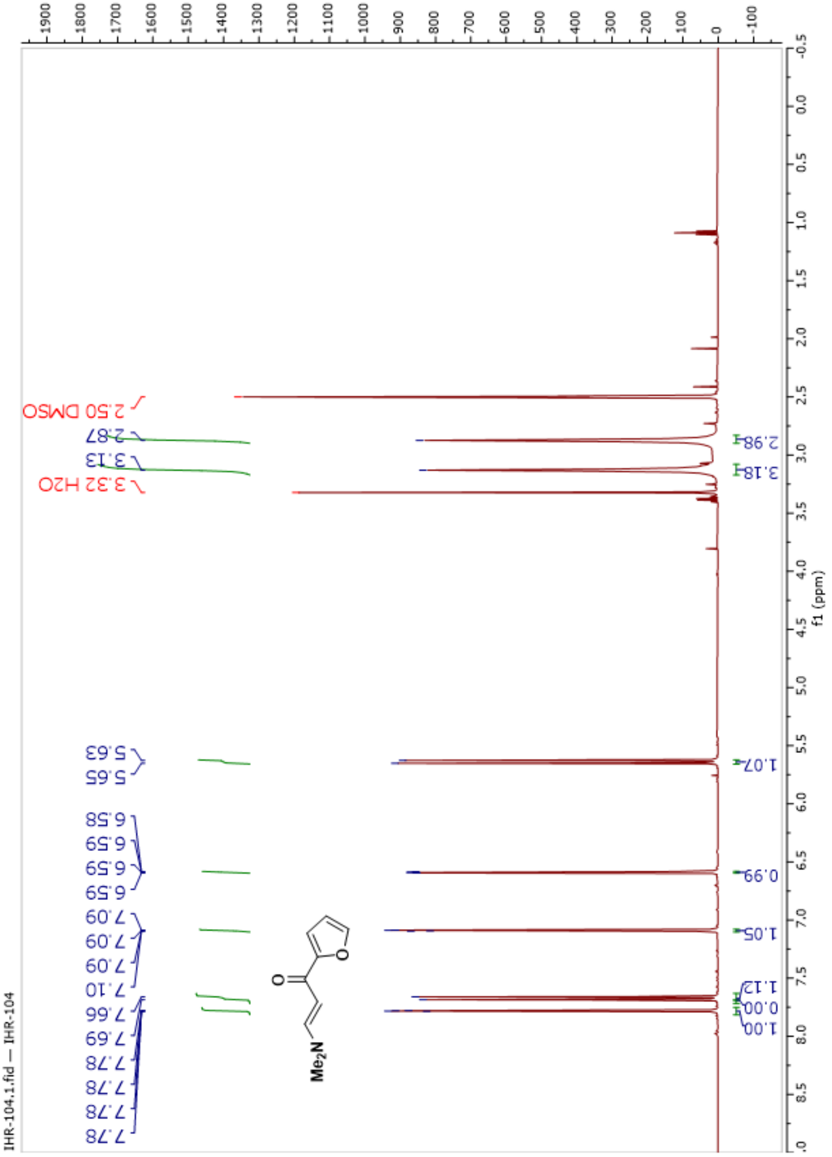
7-(Furan-2-yl)-6-(1H-indol-6-yl)pyrido[2,3-d]pyrimidine-2,4(1H,3H)-dione **(RSL-0036)**



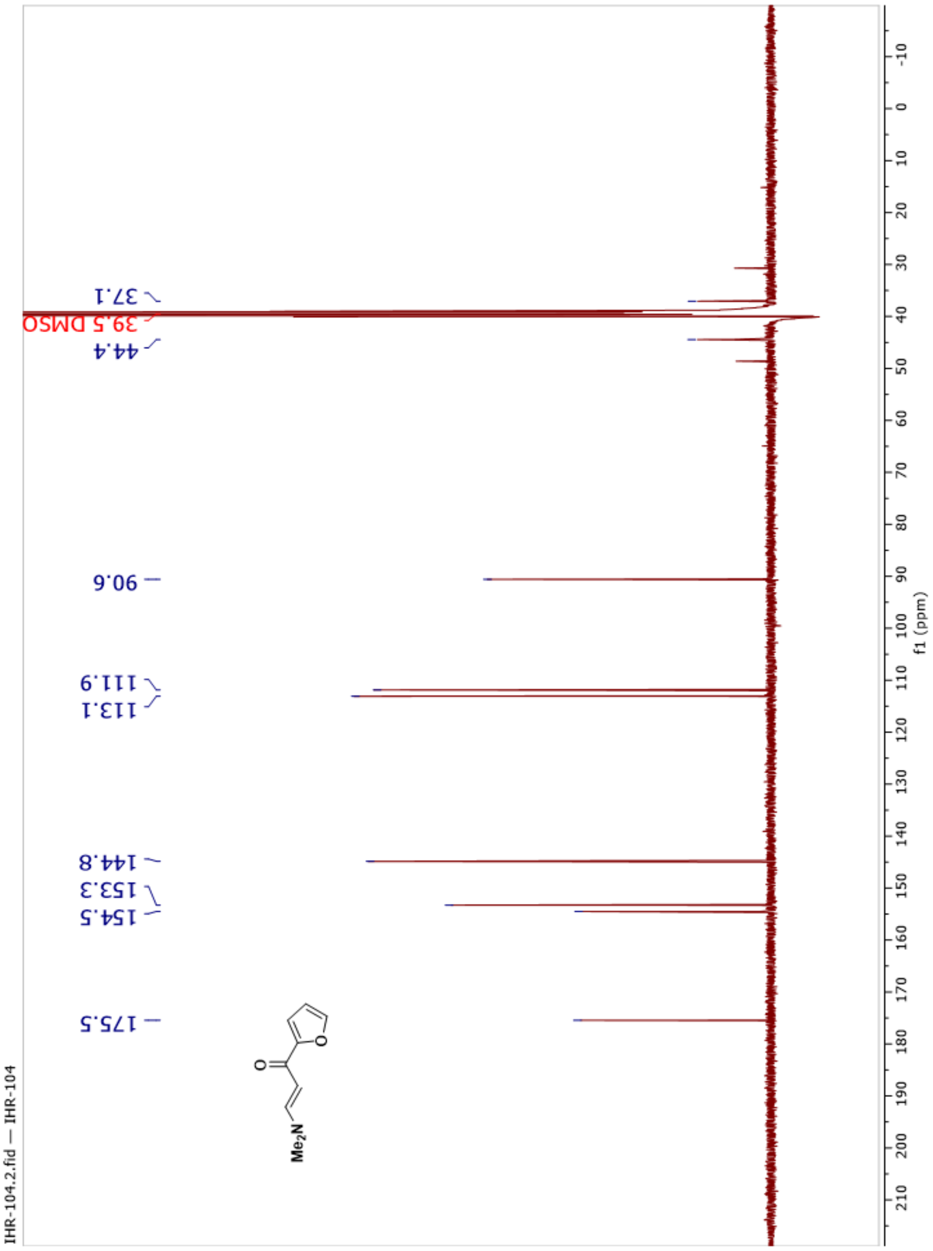
Bromide **25a** (27 mg, 0.087 mmol), Cs_2CO_3 (277 mg, 0.849 mmol) and 6-indolyl boronic acid (26 mg, 0.1615 mmol) were dissolved in 1,4-dioxane (4 mL) in a microwave tube. The tube was sealed and purged with argon and degassed under high vacuum pressure, repeated three times. $Pd(dppf)Cl_2 \cdot CH_2Cl_2$ (12 mg, 0.0352 mmol) was added to the tube, and the resulting mixture was stirred at 90 °C for 2 h. After cooling to room temperature, 1M HCl was added until pH ~7 and extracted with ethyl acetate (4 x 25 mL) and water. The combined organic layers were dried over Na_2SO_4 the crude material was then further purified using a Biotage Sfär C18 cartridge (20 μm, 12 g) eluting with acetonitrile in water, both containing 0.1% TFA (10 to 60% over 20 min, 15 mL/min). The product was eluted at 44% acetonitrile and the title compound was isolated as an orange solid (94 mg, 85%) after lyophilization.). The product was further purified on an XBridge C18 19x250mm 5 µm OBD column (H_2O/acetonitrile + 0.1% trifluoroacetic acid, 30 to 60% over 20 minutes, 15 ml/min, 370 nm, and was determined to be 98% pure.

^{1}H NMR (600 MHz, DMSO): δ = 11.80 (s, 1H), 11.45 (s, 1H), 11.21 (s, 1H), 8.01 (s, 1H), 7.75 (s, 1H), 7.61 (d, 1H, *J* = 8.0), 7.42-7.41 (m, 1H), 7.35 (s, 1H), 6.91 (dd, *J* = 8.0, 1H), 6.50 (m, 1H), 6.43-6.42 (m, 1H); ^{13}C-NMR (DMSO, 151 MHz): δ = 162.1, 151.2, 150.9, 150.5, 150.0, 144.9, 138.4, 136.0, 131.1, 130.8, 127.3, 126.3, 120.3, 120.0, 114.8, 112.0, 111.6, 107.7, 101.11; HRMS (ESI) m/z calcd for C19H13N4O3 [M+H]: 345.0988, found: 345.0982

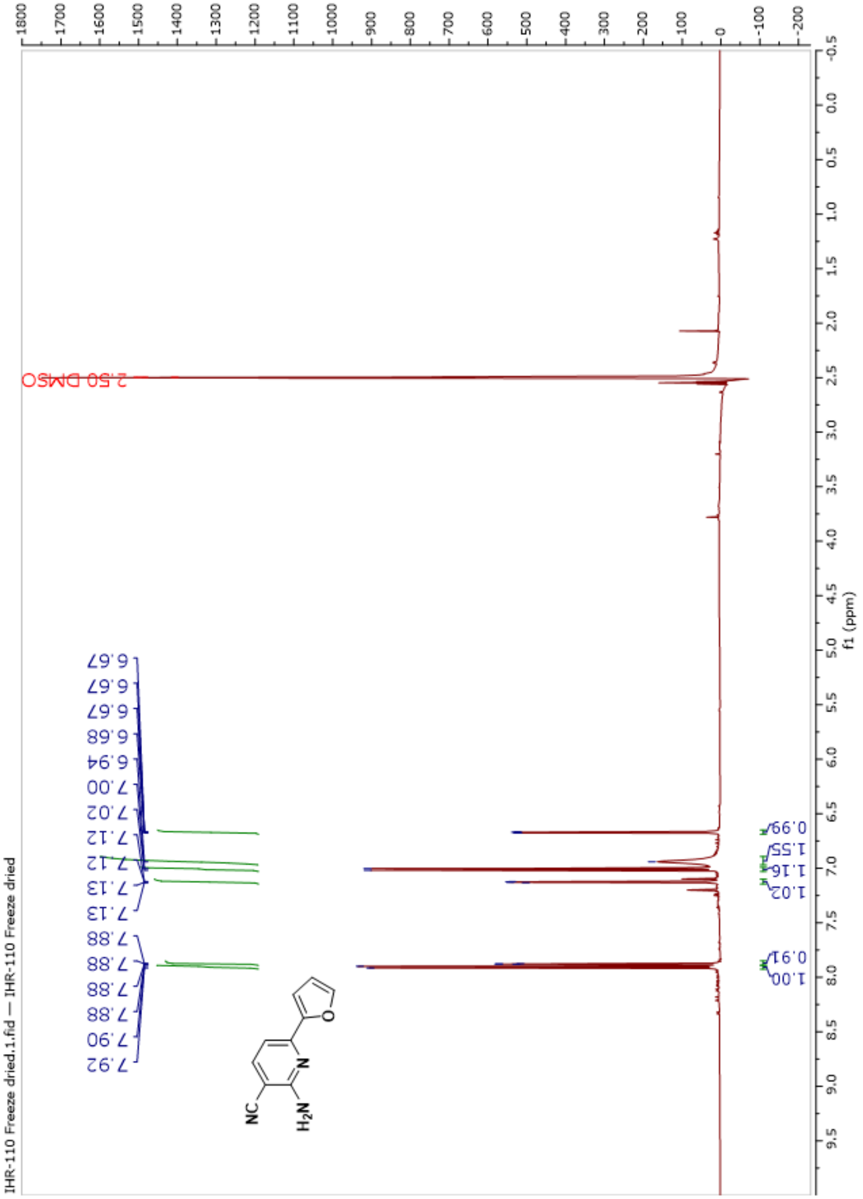
Spectrum 43: ^{1}H NMR of compound **16**



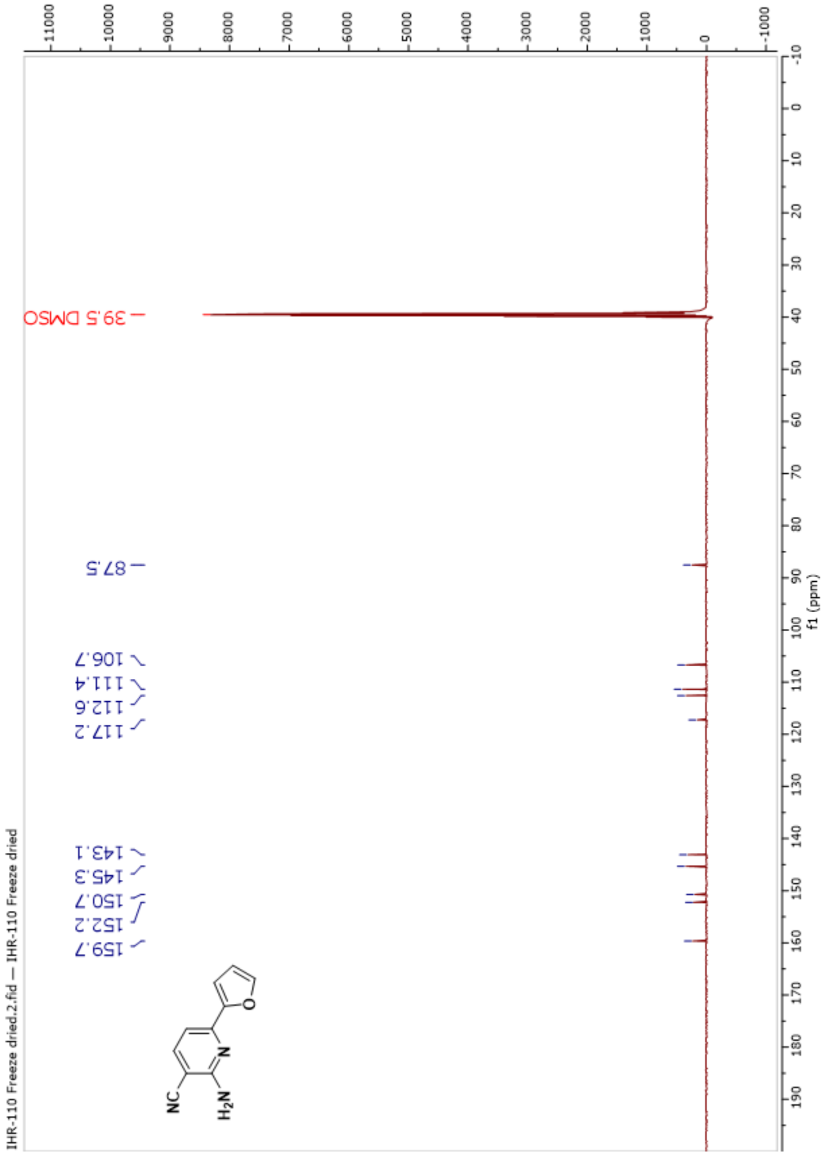
Spectrum 44: ^{13}C NMR of compound **16**



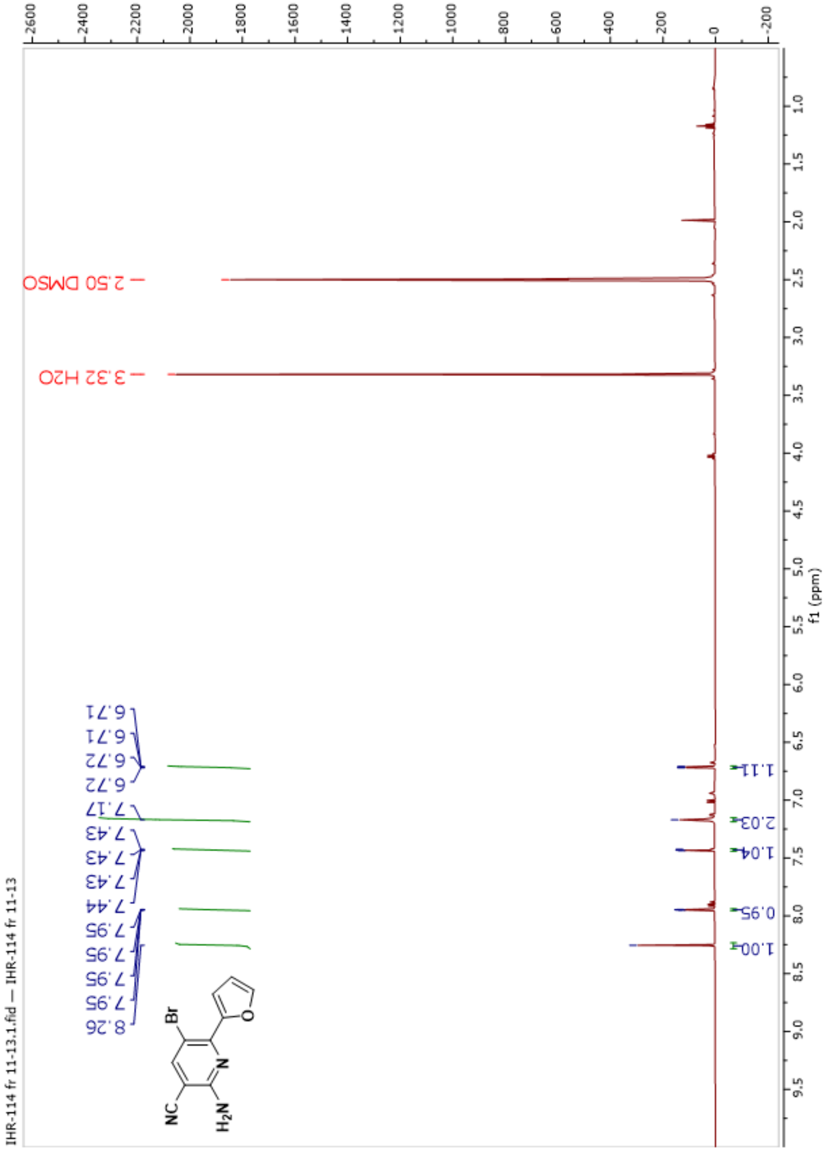
Spectrum 45: ^{1}H NMR of compound **17**



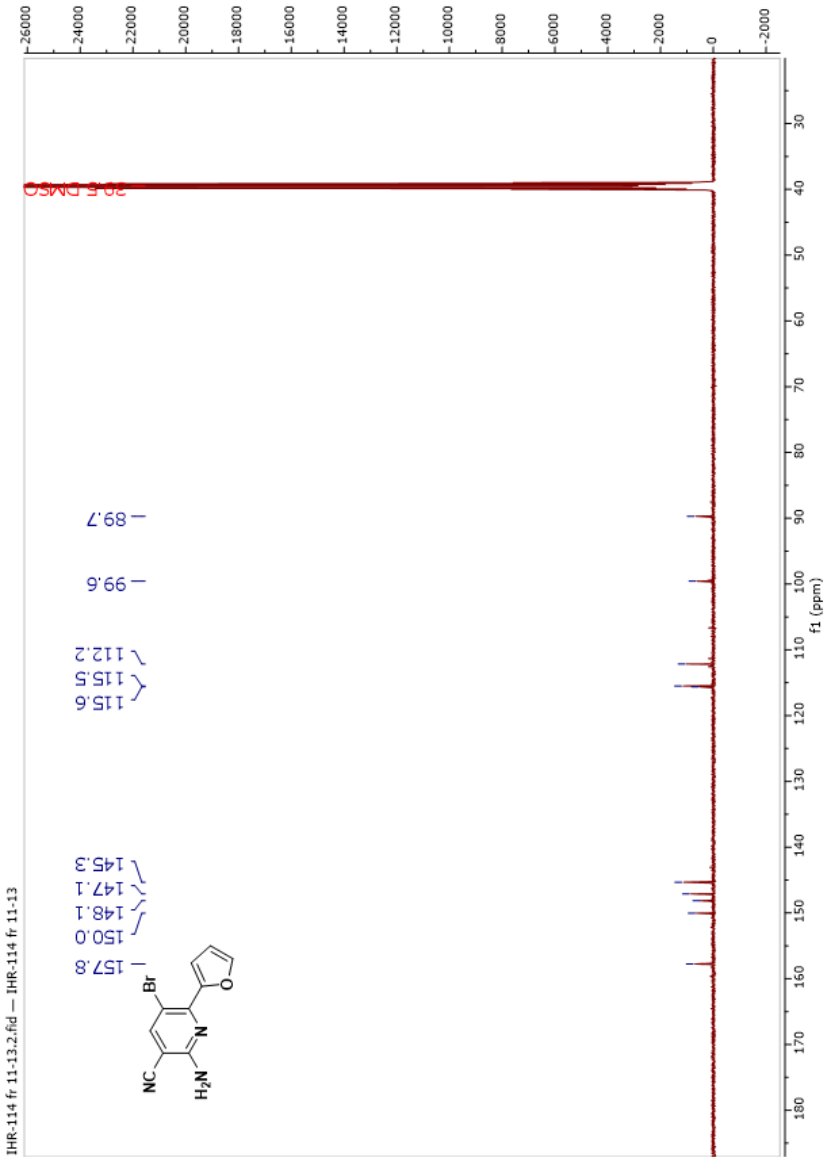
Spectrum 46: ^{13}C NMR of compound **17**



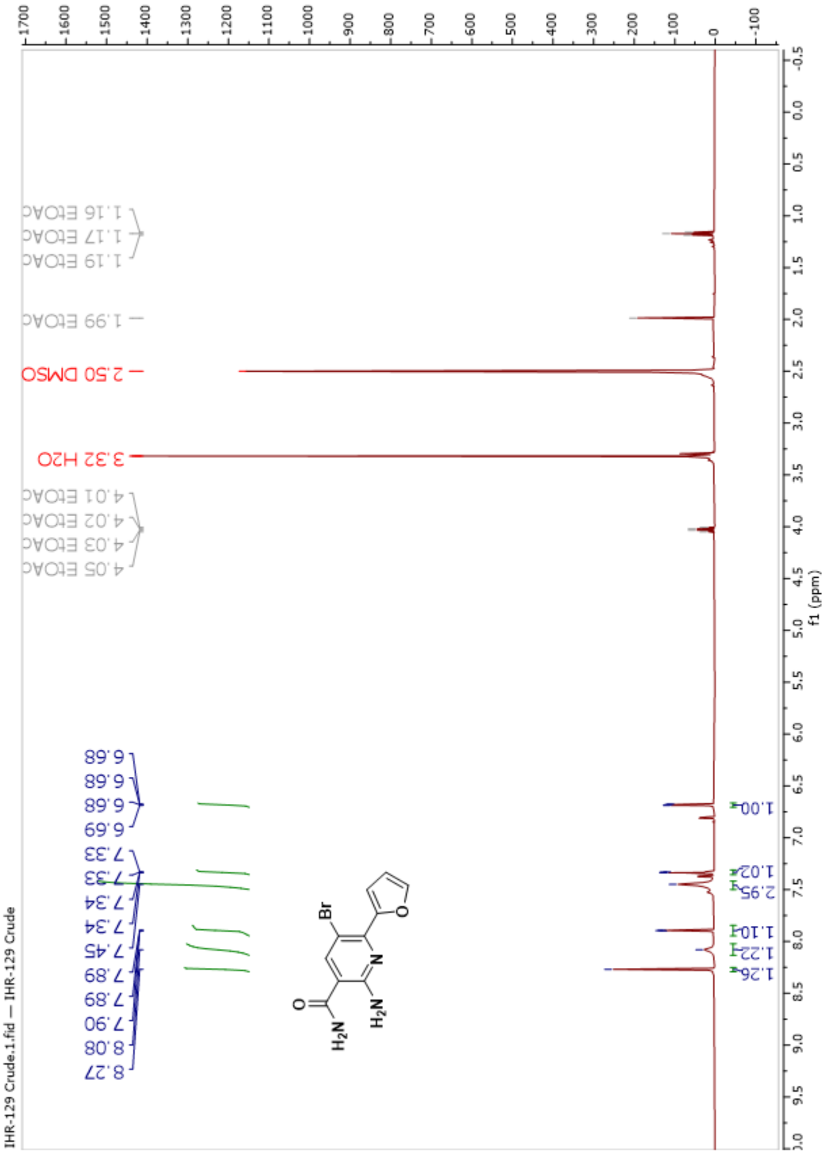
Spectrum 47: ^{1}H NMR of compound **18**



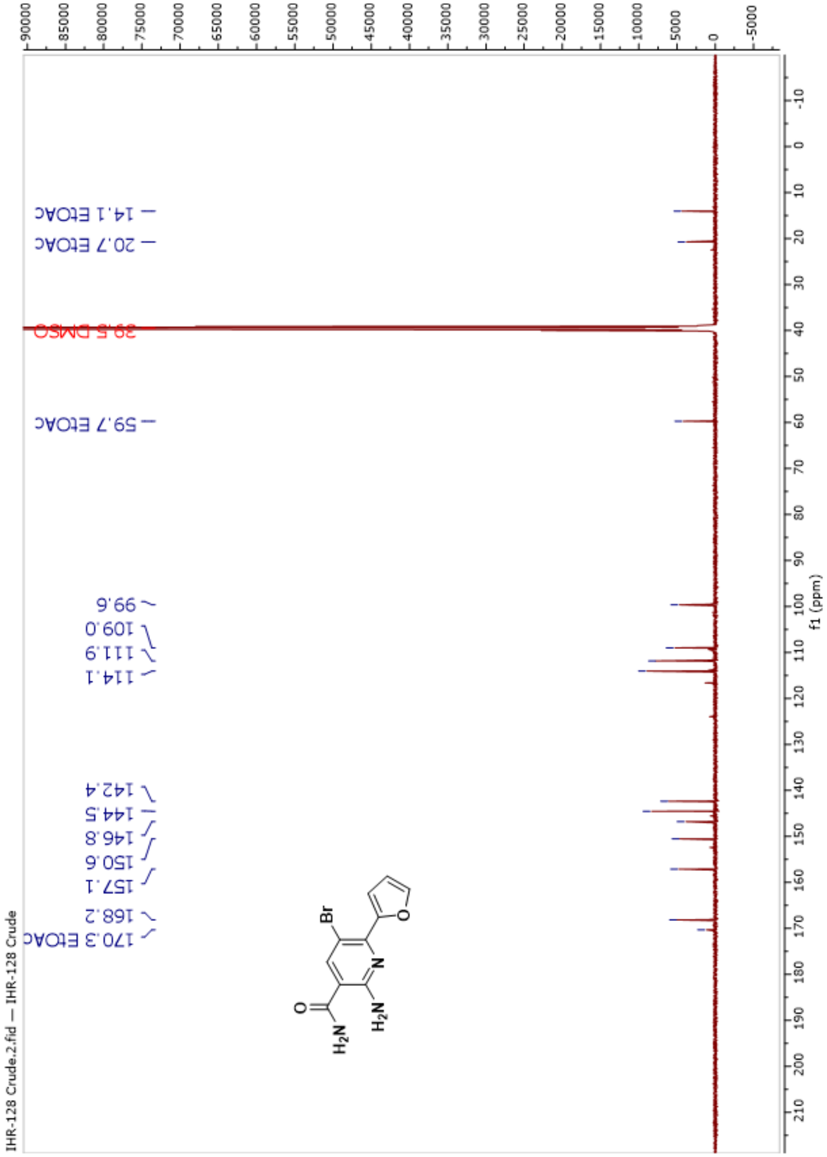
Spectrum 48: ^{13}C NMR of compound **18**



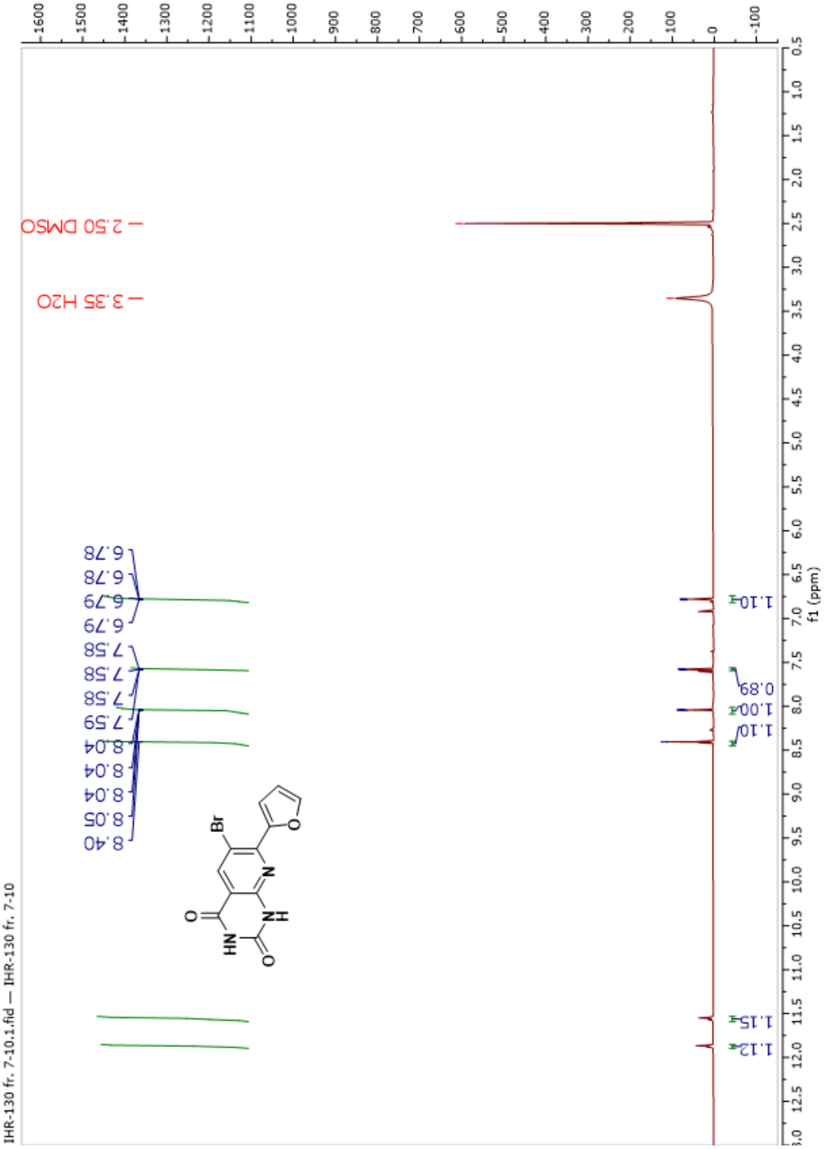
Spectrum 49: ^{1}H NMR of compound **19**



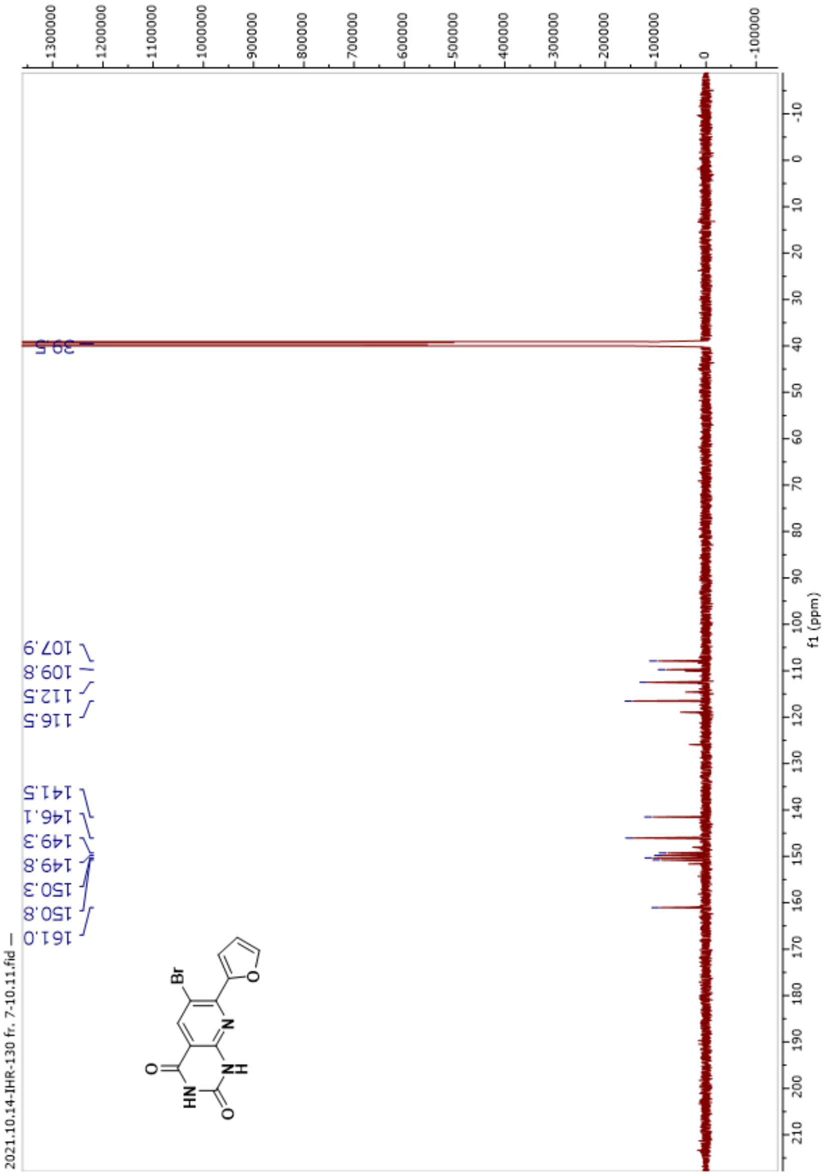
Spectrum 50: ^{13}C NMR of compound **19**



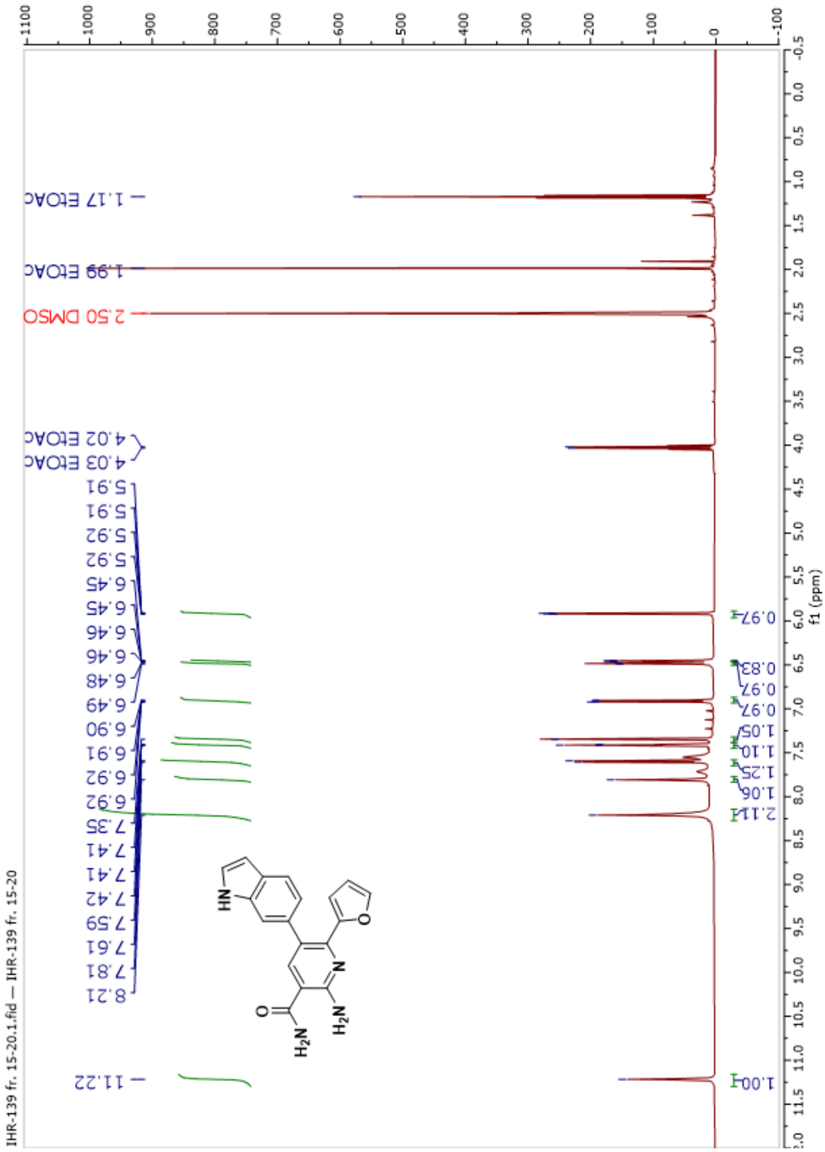
Spectrum 51: ^{1}H NMR of compound **20a**



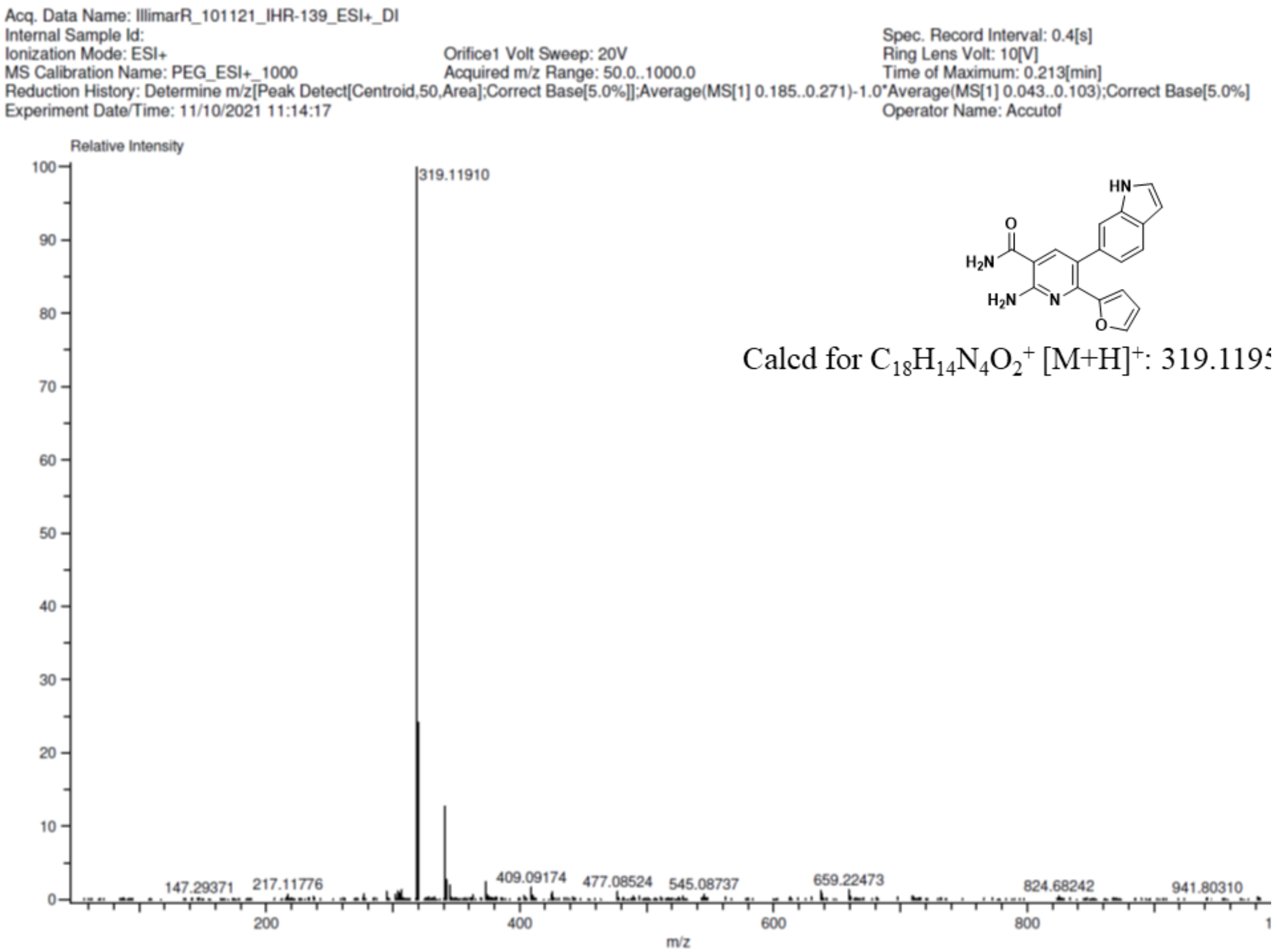
Spectrum 52: ^{13}C NMR of compound **20a**



Spectrum 53: ^{1}H NMR of compound **20b**



Spectrum 54: HRMS of compound **20b**

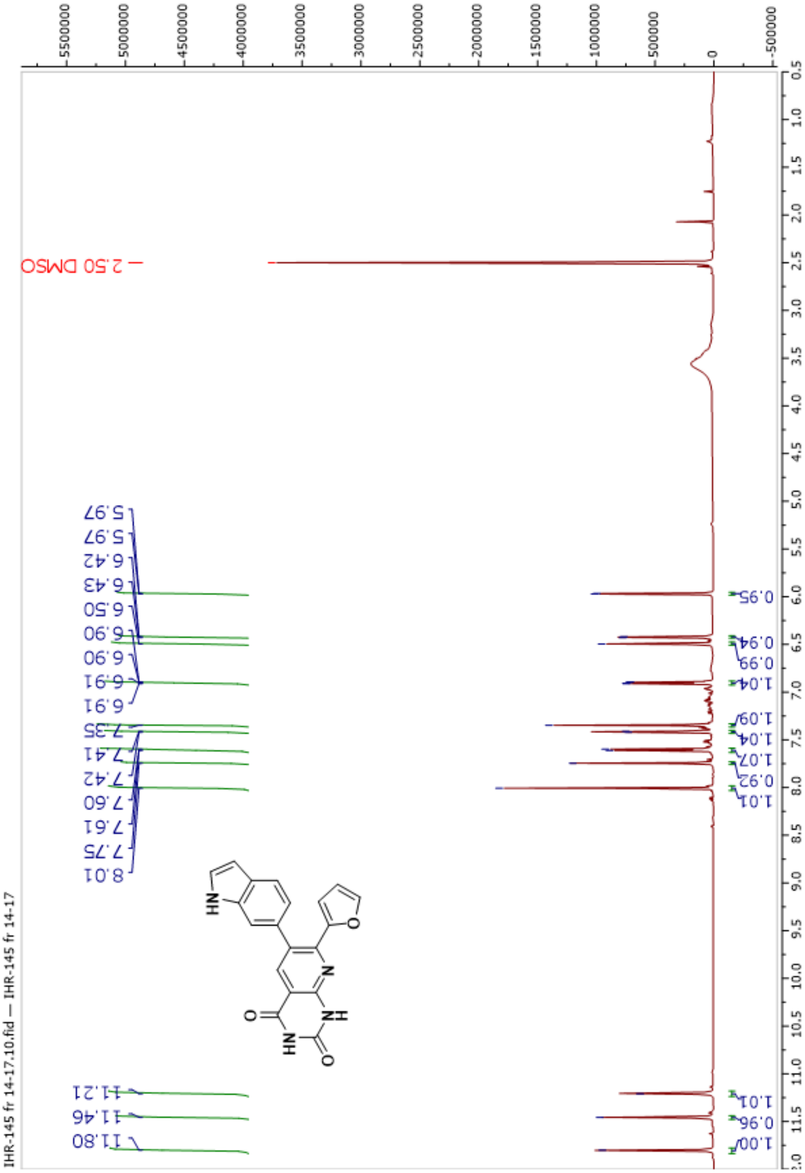


Acq. Data Name: IllimarR_101121_IHR-139_ESI+_DI
Internal Sample Id:
Ionization Mode: ESI+
MS Calibration Name: PEG_ESI+_1000
Reduction History: Determine m/z[Peak Detect[Centroid,50,Area];Correct Base[5.0%]];Average(MS[1] 0.185..0.271)-1.0*Average(MS[1] 0.043..0.103);Correct Base[5.0%]
Experiment Date/Time: 11/10/2021 11:14:17
Orifice1 Volt Sweep: 20V
Acquired m/z Range: 50.0..1000.0
Spec. Record Interval: 0.4[s]
Ring Lens Volt: 10[V]
Time of Maximum: 0.213[min]
Operator Name: Accutof

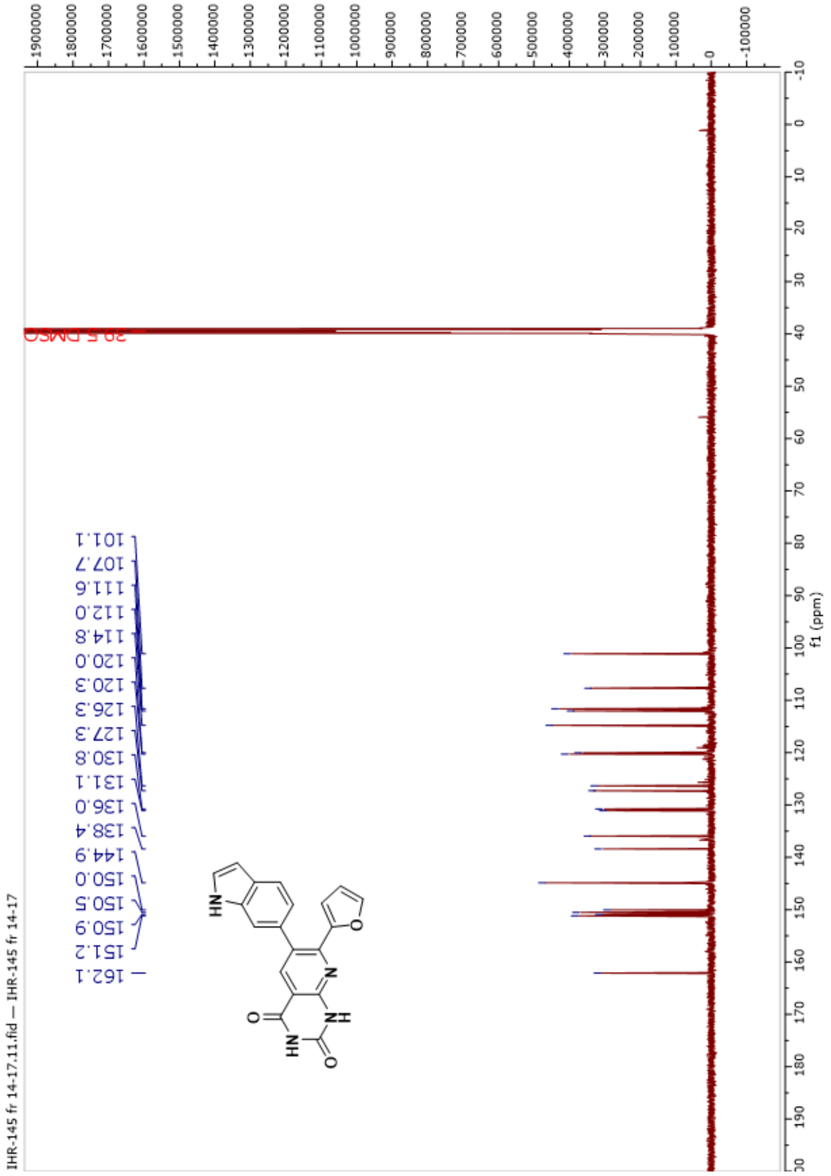
Relative Intensity
100
90
80
70
60
50
40
30
20
10
0
319.11910
147.29371
217.11776
409.09174
477.08524
545.08737
659.22473
824.68242
941.80310
200
400
600
800
1
m/z
HN
O
H$_2$N
H$_2$N
N
O

Calcd for $C_{18}H_{14}N_4O_2^+$ $[M+H]^+$: 319.1195

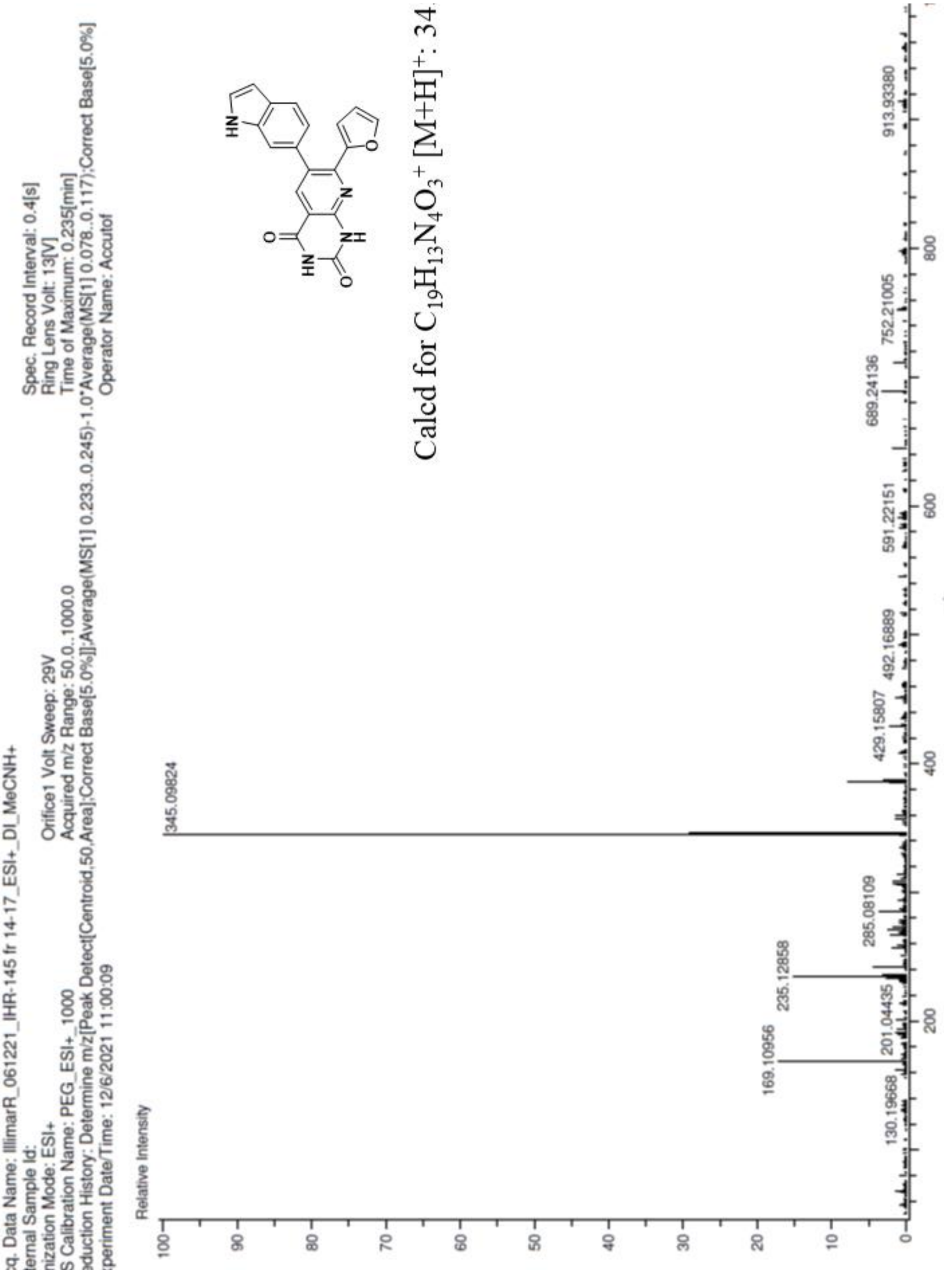
Spectrum 55: ^{1}H NMR of **RSL-0036**



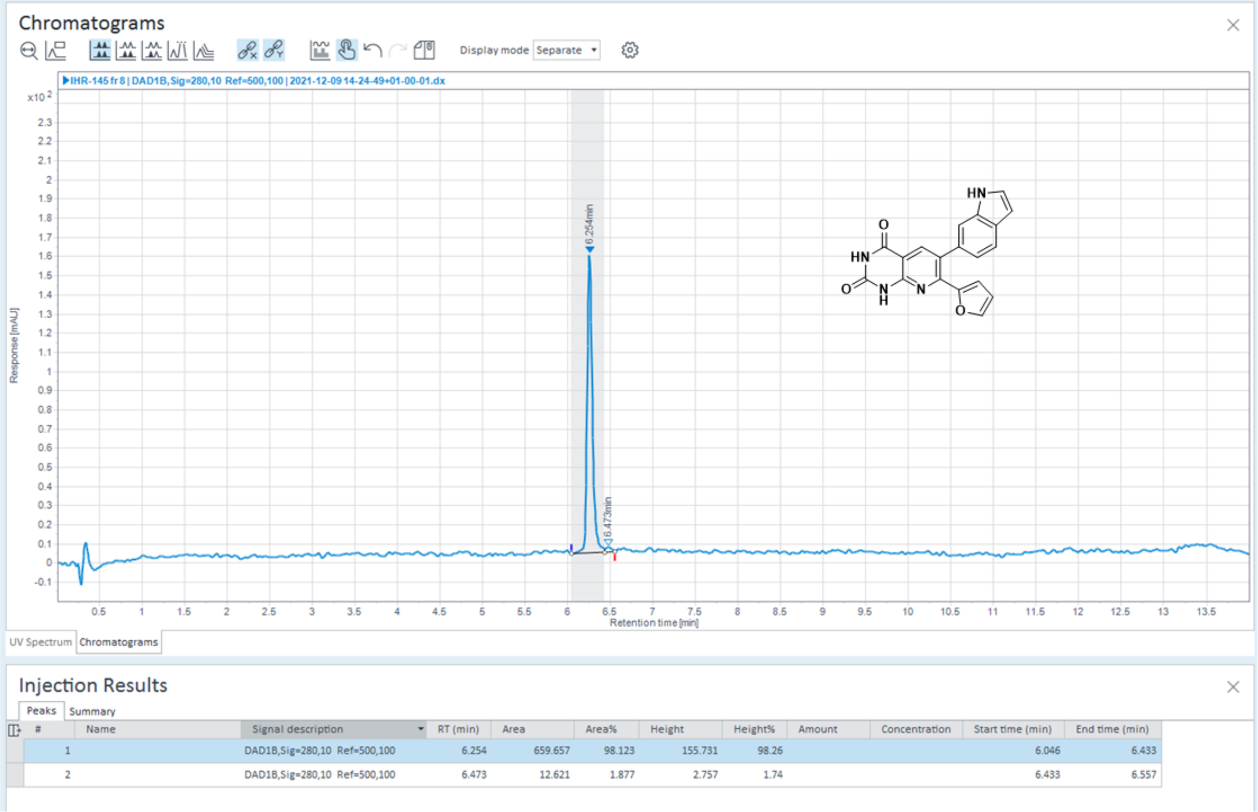
Spectrum 56: ^{13}C NMR of compound **RSL-0036**



Spectrum 57: HRMS of compound **RSL-0036**



Chromatogram 6: HPLC of **RSL-0036**


Chromatograms
Display mode Separate
▶IHR-145 fr 8 | DAD1B, Sig=280,10 Ref=500,100 | 2021-12-09 14-24-49+01-00-01.dx
x10 2
Response [mAU]
2.3
2.2
2.1
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
6.254min
6.473min
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5
9
9.5
10
10.5
11
11.5
12
12.5
13
13.5
Retention time [min]
HN
O
HN
O
N
H
N
O
UV Spectrum
Chromatograms

Injection Results

Peaks Summary

#	Name	Signal description	RT (min)	Area	Area%	Height	Height%	Amount	Concentration	Start time (min)	End time (min)
1		DAD1B,Sig=280,10 Ref=500,100	6.254	659.657	98.123	155.731	98.26			6.046	6.433
2		DAD1B,Sig=280,10 Ref=500,100	6.473	12.621	1.877	2.757	1.74			6.433	6.557

7. Appendix

Table S 1: RNA families based on overall sequence similarity. For each family, a head with PDB ID and three letter code of the small molecule bound to the pocket are listed. The total number of family members and the consensus score are also given. The druggability column contains the prediction that the majority of the members in each family obtained. (RS = riboswitch)

Family	Head	Description	Members	Consensus score	Druggability
1	1j7t_PAR	Ribosomal binding site	21	33.3	Less druggable
2	3dig_SLZ	Lysine RS	11	100.0	Less druggable
3	2eew_HPA	Guanine RS	11	100.0	Less druggable
4	3f4e_FMN	FMN RS	8	100.0	Druggable
5	3sd3_FFO	THF RS	8	100.0	Less druggable
6	6c8d_DGP	RNA-dGMP complex	5	100.0	Less druggable
7	3f2q_FMN	FMN RS	6	100.0	Less druggable
8	5nep_MGX	Guanidine RS	6	100.0	Less druggable
9	2hoj_TPP	TPP RS	7	42.86	Less druggable
10	1njn_SPS	Sparsomycin	6	100.0	Druggable
11	2gis_SAM	SAM RS	7	100.0	Druggable
12	2b57_6AP	Guanine RS	5	100.0	Less druggable
13	3owi_GLY	Glycine RS	5	100.0	Less druggable
14	4b5r_SAM	SAM RS	9	100.0	Druggable
15	2fcx_XXX	HIV-1 DIS	4	100.0	Less druggable
16	4tzx_ADE	Adenine RS	4	100.0	Less druggable
17	2cky_TPP	TPP RS	4	50.0	Druggable
18	6dlq_PRP	PRPP RS	4	50.0	Druggable

19	2ho7_G6P	glmS RZ	7	100.0	Less druggable
20	3ski_GNG	2-deoxyguanosine RS	4	100.0	Less druggable
21	3skl_GNG	2-deoxyguanosine RS	4	50.0	Less druggable
22	3irw_C2E	c-di-GMP RS	6	66.7	Druggable
23	3d0u_LYS	Lysine RS	4	100.0	Less druggable
24	6n5k_2BA	YdaO RS	3	33.3	Druggable
25	3e5c_SAM	SAM RS	3	100.0	Druggable
26	4y1i_GTP	Mn RS	3	100.0	Less druggable
27	4ts0_38E	Sprinach aptamer	4	50.0	Less druggable
28	4fel_HPA	Guanine RS	4	100.0	Less druggable
29	6e1s_HLV	PreQ1 RS	3	33.3	Druggable
30	5ddp_GLN	L-glutamine RS	2	100.0	Less druggable
31	4yaz_4BW	cGAMP RS	3	33.3	Druggable
32	3la5_5AZ	Mc6 RNA RS	3	100.0	Less druggable
33	6e8u_HZD	Mango RS	2	100.0	Druggable
34	6e1t_HLV	PreQ1 RS	2	0.0	-
35	6dmc_G4P	ppGpp RS	2	100.0	Druggable
36	3gca_PQ0	PreQ RS	2	100.0	Less duggable
37	1uud_P14	HIV-1 TAR	2	100.0	Less druggable
38	3fwo_MT9	Ribosomal subunit	2	100.0	Druggable
39	2gdi_TPP	TPP RS	2	100.0	Less druggable
40	1nta_SRY	Streptomycin	2	100.0	Less druggable
41	6c63_EKJ	Mango aptamer	2	100.0	Druggable
42	5eao_CVC	Hammerhead RZ	2	100.0	Druggable
43	4xw7_AMZ	ZMP RS	2	100.0	Less druggable

44	4qk8_2BA	c-di-AMP RS	2	100.0	Druggable
45	3bnq_PAR	Paromomycin	2	100.0	Less druggable
46	2yie_FMN	FMN RS	2	100.0	Druggable
47	3suh_FFO	THF RS	2	100.0	Less druggable
48	2o3v_N33	Ribosomal decoding site	2	100.0	Less druggable
49	6e81_TFX	Corn aptamer	2	100.0	Less druggable
50	2mxs_PAR	Neomycin RS	2	100.0	Druggable
51	2ktz_ISH	HCV IRES	2	100.0	Druggable
52	6qir_J48	CAG repeats	2	100.0	Druggable
53	6n5q_2BA	YdaO RS	2	100.0	Druggable
54	3mj3_SE4	IRES	2	100.0	Less druggable
55	6n5n_2BA	YdaO RS	2	0.0	-
56	6gzk_FH8	TMR aptamer	2	100.0	Druggable
57	1zz5_CNY	Neomycin	2	100.0	Less druggable

Table S 2: Ribosomal binding site families based on binding site sequence similarity. For each family, a head with PDB ID and three letter code of the small molecule bound to the pocket together with its name are listed. The organism associated with the binding site, the total number of family members, and the consensus score are also given. The druggability column contains the prediction that the majority of the members in each family obtained

Family	**Head**	**Ligand**	**Organism**	**Members**	**Consensus**	**Druggability**
1	1k8a_CAI	Carbomycin	*H. morismortui*	13	100.0	druggable
2	1jzx_CLY	Clindamycin	*D. radiodurans, T. thermophilus*	17	100.0	druggable
3	1fjg_PAR	Paromomycin	*T. thermophilus*	56	41.9	less druggable
4	1j5a_CTY	Clarithromycin	*D. radiodurans, T. thermophilus*	13	100	druggable
5	5jup_GDP	GDP	*S. cerevisiae*	9	77.8	less druggable
6	1ttt_GNP	GNP	*E. coli, T. aquaticus*	10	100.0	less druggable
7	4wfa_ZLD	Linezolid	*S. aureus*	7	100.0	druggable
8	6ole_MVM	PF846	*H. sapiens*	6	100.0	druggable

9	1j7t_PAR	Paromomycin	*16S rRNA Synthetic constructs*	9	11.1	druggable
10	1fjg_SRY	Spectinomycin	*T. thermophilus*	9	100.0	less druggable
11	4u3u_3HE	Cicloheximide	*S. cerevisiae*	6	100.0	druggable
12	1ibk_PAR	Paromomycin	*T. thermophilus*	30	72.4	less druggable
13	6gxm_GCP	GCP	*E. coli*	4	100.0	less druggable
14	4wpo_GDP	GDP	*T. thermophilus, E. coli*	6	100.0	less druggable
15	3jap_GCP	GCP	*K. lactis, S. cerevisiae*	4	100.0	less druggable
16	3id5_SAM	SAM/SAH	*S. solfataricus*	4	50.0	Less druggable
17	5zq0_SAH	SAH	*S. pneumoniae*	4	50.0	druggable
18	4v52_NMY	Neomycin	*E. coli, T. thermophilus*	4	0.00	-
19	6hiv_GTP	GTP	*T. brucei*	3	33.3	druggable
20	4u3m_ANM	Anisomycin	*S. cerevisiae*	8	100.0	druggable

21	1m90_SPS	Sparsomycin	*H. morismortui, S. cerevisiae*	6	66.7	druggable
22	5jup_SO1	Sordarin	*S. cerevisiae*	4	50.0	less druggable
23	6gaw_GSP	GSP	*S. scrufa, H. sapiens*	3	100.0	less druggable
24	5lzx_GCP	GCP	*H. sapiens*	3	100.0	less druggable
25	3jct_GTP	GTP	*S. cerevisiae*	4	100.0	less druggable
26	4wf1_NEG	Negamycin	*E. coli*	4	100.0	less druggable
27	1kqs_PPU	Puromycin	*H. morismortui*	3	100.0	druggable
28	6hiv_UTP	UTP	*T. brucei*	2	100.0	less druggable
29	4k32_GET	Geneticin	*Leishmania*	2	100.0	less druggable
30	4ji3_SRY	Streptomycin	*T. thermophilus*	2	100.0	less druggable
31	4io9_1F2	Carbomycin A derivative	*D. radiodurans*	2	100.0	druggable
32	1hnw_TAC	Tetracycline	*T. thermophilus*	2	100.0	less druggable
33	3wru_SJP	Neomycin analogue	-	2	100.0	less druggable

34	3td1_GET	Geneticin	-	2	100.0	less druggable
35	5lzb_GNP	GNP	*E. coli*	2	100.0	less druggable
36	5kcr_6UQ	Avilamycin	*E. coli*	2	100.0	less druggable
37	5jvg_6NO	Avilamycin	*D. radiodurans*	2	100.0	less druggable
38	5juu_SO1	Sodarin	*S. cerevisiae*	2	0.0	-
39	3oij_SAH	SAH	*S. cerevisiae*	2	100.0	less druggable
40	3jcj_GNP	GNP	*E. coli*	2	0.0	druggable
41	5aj4_GDP	GDP	*S. scrofa*	2	0.0	druggable
42	3jah_ADP	ADP	*O. corniculm, H. sapiens*	2	100.0	less druggable
43	3j7a_34G	Enetube	*P. falciparum*	2	100.0	less druggable
44	3cpw_ZLD	Linezolid	*H. morismortui*	2	100.0	druggable
45	4v9o_GCP	GCP	*T. thermophilus*	2	100.0	less druggable
46	2otj_13T	13-deoxytedanolide	*H. morismortui*	2	100.0	druggable
47	4v56_SCM	Spectinomycin	*E. coli*	2	100.0	less druggable

48	2g5k_AM2	Apramycin	*H. sapiens*	2	100.0	less druggable
49	4u56_BLS	Blasticidin S	*S. cerevisiae, T. Thermophilus*	2	100.0	less druggable
50	6rxt_GTP	GTP	*C. thermophilum*	2	100.0	less druggable
51	4u4y_PCY	Pactamycin, amicoumacin	*S. cerevisae*	2	100.0	druggable
52	6sg9_SAH	SAH	*T. brucei*	2	100.0	druggable

8. References

(1) Author Guidelines https://publish.acs.org/publish/author_guidelines?coden=jmcmar (accessed 2022 -03 -30).

(2) Armstrong, G. L.; Conn, L. A.; Pinner, R. W. Trends in Infectious Disease Mortality in the United States During the 20th Century. *JAMA* **1999**, *281* (1), 61–66. https://doi.org/10.1001/jama.281.1.61.

(3) Halliday, S. Death and Miasma in Victorian London: An Obstinate Belief. *BMJ* **2001**, *323* (7327), 1469–1471. https://doi.org/10.1136/bmj.323.7327.1469.

(4) Gest, H. The Discovery of Microorganisms by Robert Hooke and Antoni van Leeuwenhoek, Fellows of The Royal Society. *Notes Rec. R. Soc. Lond.* **2004**, *58* (2), 187–201. https://doi.org/10.1098/rsnr.2004.0055.

(5) Lister, J. XVI.—A Contribution to the Germ Theory of Putrefaction and Other Fermentative Changes, and to the Natural History of Torulæ and Bacteria. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **1875**, *27* (3), 313–344. https://doi.org/10.1017/S0080456800026004.

(6) Pasteur, L. On the Germ Theory. *Science* **1881**, *2*, 420–422. https://doi.org/DOI: 10.1126/science.os-2.62.420.

(7) Austrian, R.; Gold, J. Pneumococcal Bacteremia with Especial Reference to Bacteremic Pneumococcal Pneumonia. *Ann. Intern. Med.* **2020**. https://doi.org/doi.org/10.7326/0003-4819-60-5-759.

(8) Dineeen, P.; Homan, W. P.; Grafe, W. R. Tuberculous Peritonitis: 43 Years' Expereince in Diagnosis and Treatment. *Ann. Surg.* **1976**, *184* (6), 717–722. https://doi.org/10.1097/00000658-197612000-00010.

(9) Wainwright, M. Moulds in Folk Medicine. *Folklore* **1989**, *100* (2), 162–166.

(10) Bassett, E. J.; Keith, M. S.; Armelagos, G. J.; Martin, D. L.; Villanueva, A. R. Tetracycline-Labeled Human Bone from Ancient Sudanese Nubia (A.D. 350). *Science* **1980**. https://doi.org/10.1126/science.7001623.

(11) Williams, K. J. The Introduction of 'Chemotherapy' Using Arsphenamine – the First Magic Bullet. *J. R. Soc. Med.* **2009**, *102* (8), 343–348. https://doi.org/10.1258/jrsm.2009.09k036.

(12) Wainwright, M. On the 75th Anniversary of Prontosil. *Dyes Pigments* **2011**, *88* (3), 231–234. https://doi.org/10.1016/j.dyepig.2010.08.012.

(13) Wainwright, M. Acridine-a Neglected Antibacterial Chromophore. *J Antimicrob Chemother* **2001**, *47* (1), 1–13. https://doi.org/https://doi.org/10.1093/jac/47.1.1.

(14) Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to Their Use in the Isolation of B. Influenzæ. *Br. J. Exp. Pathol.* **1929**, *10* (3), 226–236.

(15) Robinson, F. A. Chemistry of Penicillin. *Analyst* **1947**, *72* (856), 274–276. https://doi.org/10.1039/AN9477200274.

(16) Bruyndonckx, R.; Adriaenssens, N.; Versporten, A.; Hens, N.; Monnet, D. L.; Molenberghs, G.; Goossens, H.; Weist, K.; Coenen, S.; Group, the E.-N. study; Strauss, R.; Vandael, E.; Sabtcheva, S.; Andrašević, A. T.; Kyriakidou, I.; Vlček, J.; Sönksen, U. W.; Linask, E.; Sarvikivi, E.; Hider-Mlynarz, K.; Richter, D.;

Kontopidou, F.; Matuz, M.; Aspelund, G.; Burns, K.; Fortinguerra, F.; Dimina, E.; Kuklytė, J.; Bruch, M.; Zarb, P.; Natsch, S.; Blix, H. S.; Olczak-Pieńkowska, A.; Silva, A.; Popescu, G. A.; Tesař, T.; Čižman, M.; Navas, A. L.; Bergfeldt, V.; Hopkins, S. Consumption of Antibiotics in the Community, European Union/European Economic Area, 1997–2017. *J. Antimicrob. Chemother.* **2021**, *76*, ii7-ii13. https://doi.org/10.1093/jac/dkab172.

(17) Davies, J. Where Have All the Antibiotics Gone? *Can. J. Infect. Dis. Med. Microbiol. J. Can. Mal. Infect. Microbiol. Medicale* **2006**, *17* (5), 287–290. https://doi.org/10.1155/2006/707296.

(18) Gonzalez-Estrada, A.; Radojicic, C. Penicillin Allergy: A Practical Guide for Clinicians. *Cleve. Clin. J. Med.* **2015**, *82* (5), 295. https://doi.org/10.3949/ccjm.82a.14111.

(19) Fair, R. J.; Tor, Y. Antibiotics and Bacterial Resistance in the 21st Century. *Perspect. Med. Chem.* **2014**, *6*, 25–64. https://doi.org/10.4137/PMC.S14459.

(20) Mode of Action of Chloramphenicol IX. Effects of Chloramphenicol upon a Ribosomal Amino Acid Polymerization System and Its Binding to Bacterial Ribosome. *Biochim. Biophys. Acta BBA - Nucleic Acids Protein Synth.* **1965**, *95* (1), 146–155. https://doi.org/10.1016/0005-2787(65)90219-4.

(21) Schlünzen, F.; Zarivach, R.; Harms, J.; Tocilj, A.; Albrect, T.; Yonath, A.; Franceschi, F. Structural Basis for the Interaction of Antibiotics with the Peptidyl Transferase Centre in Eubacteria. *Nature* **2001**. https://doi.org/https://doi.org/10.1038/35101544.

(22) Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakhrisnan, V. Functional Insights from the Structure of the 30S Ribosomal Subunit and Its Interactions with Antibiotics. *Nature* **2000**. https://doi.org/https://doi.org/10.1038/35030019.

(23) Anokhina, M. M.; Barta, A.; Nierhaus, K. H.; Spiridonova, V. A.; Kopylov, A. M. Mapping of the Second Tetracycline Binding Site on the Ribosomal Small Subunit of E.Coli. *Nucleic Acids Res.* **2004**, *32* (8), 2594–2597. https://doi.org/10.1093/nar/gkh583.

(24) Hammes, W. P.; Neuhaus, F. C. On the Mechanism of Action of Vancomycin: Inhibition of Peptidoglycan Synthesis in Gaffkya Homari. *Antimicrob. Agents Chemother.* **1974**. https://doi.org/10.1128/AAC.6.6.722.

(25) Campbell, E. A.; Korzheva, N.; Mustaev, A.; Murakami, K.; Nair, S.; Goldfarb, A.; Darst, S. A. Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* **2001**, *104* (6), 901–912. https://doi.org/https://doi.org/10.1016/S0092-8674(01)00286-0.

(26) Aldred, K. J.; Kerns, R. J.; Osheroff, N. Mechanism of Quinolone Action and Resistance. *Biochemistry (Mosc.)* **2014**, *53* (10), 1565–1574. https://doi.org/10.1021/bi5000564.

(27) Fang, S.-G.; Qiang, T.; Liu, R.-J.; Xu, X.-M.; Zhang, Y.-W. Enhanced Production of 6-Aminopenicillanic Acid in Aqueous Methyl Isobutyl Ketone System with Immobilized Penicillin G Acylase. *Prep. Biochem. Biotechnol.* **2009**, *40* (1), 38–45. https://doi.org/10.1080/10826060903389489.

(28) Maresova, H.; Plačková, M.; Grulich, M.; Kyslík, P. Current State and Perspectives of Penicillin G Acylase-Based Biocatalyses. *Appl. Microbiol. Biotechnol.* **2014**, *98* (7), 2867–2879. https://doi.org/10.1007/s00253-013-5492-7.

(29) Davies, J.; Davies, D. Origins and Evolution of Antibiotic Resistance. *Microbiol. Mol. Biol. Rev. MMBR* **2010**, *74* (3), 417–433. https://doi.org/10.1128/MMBR.00016-10.

(30) Klevens, R. M.; Morrison, M. A.; Nadle, J.; Petit, S.; Gershman, K.; Ray, S.; Harrison, L. H.; Lynfield, R.; Dumyati, G.; Townes, J. M.; Craig, A. S.; Zell, E. R.; Fosheim, G. E.; McDougal, L. K.; Carey, R. B.; Fridkin, S. K.; Active Bacterial Core surveillance (ABCs) MRSA Investigators, for the. Invasive Methicillin-Resistant Staphylococcus Aureus Infections in the United States. *JAMA* **2007**, *298* (15), 1763–1771. https://doi.org/10.1001/jama.298.15.1763.

(31) Cassini, A.; Högberg, L. D.; Plachouras, D.; Quattrocchi, A.; Hoxha, A.; Simonsen, G. S.; Colomb-Cotinat, M.; Kretzschmar, M. E.; Devleesschauwer, B.; Cecchini, M.; Ouakrim, D. A.; Oliveira, T. C.; Struelens, M. J.; Suetens, C.; Monnet, D. L.; Strauss, R.; Mertens, K.; Struyf, T.; Catry, B.; Latour, K.; Ivanov, I. N.; Dobreva, E. G.; Andrasevic, A. T.; Soprek, S.; Budimir, A.; Paphitou, N.; Žemlicková, H.; Olsen, S. S.; Sönksen, U. W.; Märtin, P.; Ivanova, M.; Lyytikäinen, O.; Jalava, J.; Coignard, B.; Eckmanns, T.; Sin, M. A.; Haller, S.; Daikos, G. L.; Gikas, A.; Tsiodras, S.; Kontopidou, F.; Tóth, Á.; Hajdu, Á.; Guólaugsson, Ó.; Kristinsson, K. G.; Murchan, S.; Burns, K.; Pezzotti, P.; Gagliotti, C.; Dumpis, U.; Liuimiene, A.; Perrin, M.; Borg, M. A.; Greeff, S. C. de; Monen, J. C.; Koek, M. B.; Elstrøm, P.; Zabicka, D.; Deptula, A.; Hryniewicz, W.; Caniça, M.; Nogueira, P. J.; Fernandes, P. A.; Manageiro, V.; Popescu, G. A.; Serban, R. I.; Schréterová, E.; Litvová, S.; Stefkovica, M.; Kolman, J.; Klavs, I.; Korošec, A.; Aracil, B.; Asensio, A.; Pérez-Vázquez, M.; Billström, H.; Larsson, S.; Reilly, J. S.; Johnson, A.; Hopkins, S. Attributable Deaths and Disability-Adjusted Life-Years Caused by Infections with Antibiotic-Resistant Bacteria in the EU and the European Economic Area in 2015: A Population-Level Modelling Analysis. *Lancet Infect. Dis.* **2019**, *19* (1), 56–66. https://doi.org/10.1016/S1473-3099(18)30605-4.

(32) Murray, C. J.; Ikuta, K. S.; Sharara, F.; Swetschinski, L.; Aguilar, G. R.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; Johnson, S. C.; Browne, A. J.; Chipeta, M. G.; Fell, F.; Hackett, S.; Haines-Woodhouse, G.; Hamadani, B. H. K.; Kumaran, E. A. P.; McManigal, B.; Agarwal, R.; Akech, S.; Albertson, S.; Amuasi, J.; Andrews, J.; Aravkin, A.; Ashley, E.; Bailey, F.; Baker, S.; Basnyat, B.; Bekker, A.; Bender, R.; Bethou, A.; Bielicki, J.; Boonkasidecha, S.; Bukosia, J.; Carvalheiro, C.; Castañeda-Orjuela, C.; Chansamouth, V.; Chaurasia, S.; Chiurchiù, S.; Chowdhury, F.; Cook, A. J.; Cooper, B.; Cressey, T. R.; Criollo-Mora, E.; Cunningham, M.; Darboe, S.; Day, N. P. J.; Luca, M. D.; Dokova, K.; Dramowski, A.; Dunachie, S. J.; Eckmanns, T.; Eibach, D.; Emami, A.; Feasey, N.; Fisher-Pearson, N.; Forrest, K.; Garrett, D.; Gastmeier, P.; Giref, A. Z.; Greer, R. C.; Gupta, V.; Haller, S.; Haselbeck, A.; Hay, S. I.; Holm, M.; Hopkins, S.; Iregbu, K. C.; Jacobs, J.; Jarovsky, D.; Javanmardi, F.; Khorana, M.; Kissoon, N.; Kobeissi, E.; Kostyanev, T.; Krapp, F.; Krumkamp, R.; Kumar, A.; Kyu, H. H.; Lim, C.; Limmathurotsakul, D.; Loftus, M. J.; Lunn, M.; Ma, J.; Mturi, N.; Munera-Huertas, T.; Musicha, P.; Mussi-Pinhata, M. M.; Nakamura, T.; Nanavati, R.; Nangia, S.; Newton, P.; Ngoun, C.; Novotney, A.; Nwakanma, D.;

Obiero, C. W.; Olivas-Martinez, A.; Olliaro, P.; Ooko, E.; Ortiz-Brizuela, E.; Peleg, A. Y.; Perrone, C.; Plakkal, N.; Ponce-de-Leon, A.; Raad, M.; Ramdin, T.; Riddell, A.; Roberts, T.; Robotham, J. V.; Roca, A.; Rudd, K. E.; Russell, N.; Schnall, J.; Scott, J. A. G.; Shivamallappa, M.; Sifuentes-Osornio, J.; Steenkeste, N.; Stewardson, A. J.; Stoeva, T.; Tasak, N.; Thaiprakong, A.; Thwaites, G.; Turner, C.; Turner, P.; Doorn, H. R. van; Velaphi, S.; Vongpradith, A.; Vu, H.; Walsh, T.; Waner, S.; Wangrangsimakul, T.; Wozniak, T.; Zheng, P.; Sartorius, B.; Lopez, A. D.; Stergachis, A.; Moore, C.; Dolecek, C.; Naghavi, M. Global Burden of Bacterial Antimicrobial Resistance in 2019: A Systematic Analysis. *The Lancet* **2022**, *399* (10325), 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0.

(33) Rice, L. B. The Maxwell Finland Lecture: For the Duration— Rational Antibiotic Administration in an Era of Antimicrobial Resistance and Clostridium Difficile. *Clin. Infect. Dis.* **2008**, *46* (4), 491–496. https://doi.org/10.1086/526535.

(34) Palumbi, S. R. Humans as the World's Greatest Evolutionary Force. *Science* **2001**. https://doi.org/10.1126/science.293.5536.1786.

(35) Mead, P. S.; Slutsker, L.; Dietz, V.; McCaig, L. F.; Bresee, J. S.; Shapiro, C.; Griffin, P. M.; Tauxe, R. V. Food-Related Illness and Death in the United States. *Emerg. Infect. Dis.* **1999**, *5* (5), 607–625. https://doi.org/10.3201/eid0505.990502.

(36) Lyon, B. R.; Skurray, R. Antimicrobial Resistance of Staphylococcus Aureus: Genetic Basis. *Microbiol. Rev.* **1987**, *51* (1), 88–134.

(37) Schaack, S.; Gilbert, C.; Feschotte, C. Promiscuous DNA: Horizontal Transfer of Transposable Elements and Why It Matters for Eukaryotic Evolution. *Trends Ecol. Evol.* **2010**, *25* (9), 537–546. https://doi.org/10.1016/j.tree.2010.06.001.

(38) Hall, B. G.; Barlow, M. Structure-Based Phylogenies of the Serine β-Lactamases. *J. Mol. Evol.* **2003**, *57* (3), 255–260. https://doi.org/10.1007/s00239-003-2473-y.

(39) Fleming, S. A. The Nobel Prize in Physiology or Medicine 1945 https://www.nobelprize.org/prizes/medicine/1945/fleming/lecture/ (accessed 2021 -07 -12).

(40) Fischbach, M. A.; Walsh, C. T. Antibiotics For Emerging Pathogens. *Science* **2009**, *325* (5944), 1089–1093. https://doi.org/10.1126/science.1176667.

(41) Shah, N. S.; Wright, A.; Bai, G.-H.; Barrera, L.; Boulahbal, F.; Martín-Casabona, N.; Drobniewski, F.; Gilpin, C.; Havelková, M.; Lepe, R.; Lumb, R.; Metchock, B.; Portaels, F.; Rodrigues, M. F.; Rüsch-Gerdes, S.; Van Deun, A.; Vincent, V.; Laserson, K.; Wells, C.; Cegielski, J. P. Worldwide Emergence of Extensively Drug-Resistant Tuberculosis. *Emerg. Infect. Dis.* **2007**, *13* (3), 380–387. https://doi.org/10.3201/eid1303.061400.

(42) Global action plan on antimicrobial resistance https://www.who.int/publications-detail-redirect/9789241509763 (accessed 2021 -10 -02).

(43) WHO publishes list of bacteria for which new antibiotics are urgently needed https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed (accessed 2022 -03 -27).

(44) O'Neill, J. *Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations*; Wellcome Collection, 2014.

(45) Infectious Diseases Society of America. The 10 x '20 Initiative: Pursuing a Global Commitment to Develop 10 New Antibacterial Drugs by 2020. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **2010**, *50* (8), 1081–1083. https://doi.org/10.1086/652237.

(46) Talbot, G. H.; Jezek, A.; Murray, B. E.; Jones, R. N.; Ebright, R. H.; Nau, G. J.; Rodvold, K. A.; Newland, J. G.; Boucher, H. W.; Infectious Diseases Society of America. The Infectious Diseases Society of America's 10 × '20 Initiative (10 New Systemic Antibacterial Agents US Food and Drug Administration Approved by 2020): Is 20 × '20 a Possibility? *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **2019**, *69* (1), 1–11. https://doi.org/10.1093/cid/ciz089.

(47) Hay, M.; Thomas, D. W.; Craighead, J. L.; Economides, C.; Rosenthal, J. Clinical Development Success Rates for Investigational Drugs. *Nat. Biotechnol.* **2014**, *32* (1), 40–51. https://doi.org/10.1038/nbt.2786.

(48) Munos, B. Lessons from 60 Years of Pharmaceutical Innovation. *Nat. Rev. Drug Discov.* **2009**, *8* (12), 959–968. https://doi.org/10.1038/nrd2961.

(49) Hingorani, A. D.; Kuan, V.; Finan, C.; Kruger, F. A.; Gaulton, A.; Chopade, S.; Sofat, R.; MacAllister, R. J.; Overington, J. P.; Hemingway, H.; Denaxas, S.; Prieto, D.; Casas, J. P. Improving the Odds of Drug Development Success through Human Genomics: Modelling Study. *Sci. Rep.* **2019**, *9* (1), 1–25. https://doi.org/10.1038/s41598-019-54849-w.

(50) Brown, D.; Superti-Furga, G. Rediscovering the Sweet Spot in Drug Discovery. *Drug Discov. Today* **2003**, *8* (23), 1067–1077. https://doi.org/10.1016/S1359-6446(03)02902-7.

(51) Bains, W. Failure Rates in Drug Discovery and Development. *Drug Discov.* **2004**, 9.

(52) Edfeldt, F. N.; Folmer, R. H.; Breeze, A. L. Fragment Screening to Predict Druggability (Ligandability) and Lead Discovery Success. *Drug Discov. Today* **2011**. https://doi.org/https://doi.org/10.1016/j.drudis.2011.02.002.

(53) Abi Hussein, H.; Geneix, C.; Petitjean, M.; Borrel, A.; Flatters, D.; Camproux, A.-C. Global Vision of Druggability Issues: Applications and Perspectives. *Drug Discov. Today* **2017**, *22* (2), 404–415. https://doi.org/10.1016/j.drudis.2016.11.021.

(54) Hopkins, A. L.; Groom, C. R. The Druggable Genome. *Nat. Rev. Drug Discov.* **2002**, *1*, 727–730. https://doi.org/https://doi.org/10.1038/nrd892.

(55) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How Many Drug Targets Are There? *Nat. Rev. Drug Discov.* **2006**, *5* (12), 993–996. https://doi.org/10.1038/nrd2199.

(56) Kubinyi, H. Drug Research: Myths, Hype and Reality. *Nat. Rev. Drug Discov.* **2003**, *2* (8), 665–668. https://doi.org/10.1038/nrd1156.

(57) Schmidtke, P.; Barril, X. Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *J. Med. Chem.* **2010**, *53* (15), 5858–5867. https://doi.org/10.1021/jm100574m.

(58) Macarron, R. Critical Review of the Role of HTS in Drug Discovery. *Drug Discov. Today* **2006**, *11* (7–8), 277–279. https://doi.org/10.1016/j.drudis.2006.02.001.

(59) Gupta, A.; Gupta, A. K.; Seshadri, K. Structural Models in the Assessment of Protein Druggability Based on HTS Data. *J. Comput. Aided Mol. Des.* **2009**, *23* (8), 583–592. https://doi.org/10.1007/s10822-009-9279-y.

(60) Ertl, P. Cheminformatics Analysis of Organic Substituents: Identification of the Most Common Substituents, Calculation of Substituent Properties, and Automatic Identification of Drug-like Bioisosteric Groups. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (2), 374–380. https://doi.org/10.1021/ci0255782.

(61) Erlanson, D. A.; Wells, J. A.; Braisted, A. C. Tethering: Fragment-Based Drug Discovery. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33* (1), 199–223. https://doi.org/10.1146/annurev.biophys.33.110502.140409.

(62) Chen, I.-J.; Hubbard, R. E. Lessons for Fragment Library Design: Analysis of Output from Multiple Screening Campaigns. *J. Comput. Aided Mol. Des.* **2009**, *23* (8), 603–620. https://doi.org/10.1007/s10822-009-9280-5.

(63) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. Druggability Indices for Protein Targets Derived from NMR-Based Screening Data. *J. Med. Chem.* **2005**, *48* (7), 2518–2525. https://doi.org/10.1021/jm049131r.

(64) Fragment-Based Target Screening as an Empirical Approach to Prioritising Targets: A Case Study on Antibacterials. *Drug Discov. Today* **2020**, *25* (11), 2030–2037. https://doi.org/10.1016/j.drudis.2020.09.003.

(65) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009**, *49* (2), 377–389. https://doi.org/10.1021/ci800324m.

(66) Seco, J.; Luque, F. J.; Barril, X. Binding Site Detection and Druggability Index from First Principles. *J. Med. Chem.* **2009**, *52* (8), 2363–2371. https://doi.org/10.1021/jm801385d.

(67) Hajduk, P. J.; Huth, J. R.; Tse, C. Predicting Protein Druggability. *Drug Discov. Today* **2005**, *10* (23–24), 1675–1682. https://doi.org/10.1016/S1359-6446(05)03624-X.

(68) Henrich, S.; Salo-Ahen, O. M. H.; Huang, B.; Rippmann, F. F.; Cruciani, G.; Wade, R. C. Computational Approaches to Identifying and Characterizing Protein Binding Sites for Ligand Design. *J. Mol. Recognit.* **2010**, *23* (2), 209–219. https://doi.org/10.1002/jmr.984.

(69) Fauman, E. B.; Kai, B. K.; Huang, E. S. Structure-Based Druggability Assessment — Identifying Suitable Targets for Small Molecule Therapeutics. *Curr. Opin. Chem. Biol.* **2011**, *15* (4), 463–468. https://doi.org/10.1016/j.cbpa.2011.05.020.

(70) Kandoi, G.; Acencio, M. L.; Lemke, N. Prediction of Druggable Proteins Using Machine Learning and Systems Biology: A Mini-Review. *Front. Physiol.* **2015**, *0*. https://doi.org/10.3389/fphys.2015.00366.

(71) Krasowski, A.; Muthas, D.; Sarkar, A.; Schmitt, S.; Brenk, R. DrugPred: A Structure-Based Approach to Predict Protein Druggability Developed Using an Extensive Nonredundant Data Set. *J. Chem. Inf. Model.* **2011**, *51* (11), 2829–2842. https://doi.org/https://doi.org/10.1021/ci200266d.

(72) Sarkar, A.; Brenk, R. To Hit or Not to Hit, That Is the Question – Genome-Wide Structure-Based Druggability Predictions for Pseudomonas Aeruginosa Proteins. *PLOS ONE* **2015**, *10* (9), e0137279. https://doi.org/10.1371/journal.pone.0137279.

(73) Imming, P.; Sinning, C.; Meyer, A. Drugs, Their Targets and the Nature and Number of Drug Targets. *Nat. Rev. Drug Discov.* **2006**, *5* (10), 821–834. https://doi.org/https://doi.org/10.1038/nrd2132.

(74) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. A Comprehensive Map of Molecular Drug Targets. *Nat. Rev. Drug Discov.* **2017**, *16* (1), 19–34. https://doi.org/10.1038/nrd.2016.230.

(75) Maccubbin, A. E.; Caballes, L.; Scappaticci, F.; Struck, R. F.; Gurtoo, H. L. 32P-Postlabeling Analysis of Binding of the Cyclophosphamide Metabolite, Acrolein, to DNA. *Cancer Commun.* **1990**. https://doi.org/10.3727/095535490820874380.

(76) Temperini, C.; Messori, L.; Orioli, P.; Bugno, C.; Animati, F.; Ughetto, G. The Crystal Structure of the Complex between a Disaccharide Anthracycline and the DNA Hexamer d(CGATCG) Reveals Two Different Binding Sites Involving Two DNA Duplexes. *Nucleic Acids Res.* **2003**. https://doi.org/10.1093/nar/gkg245.

(77) Wilson, D. N. Ribosome-Targeting Antibiotics and Mechanisms of Bacterial Resistance. *Nat. Rev. Microbiol.* **2013**, *12* (1), 35–48. https://doi.org/10.1038/nrmicro3155.

(78) Vicens, Q.; Westhof, E. RNA as a Drug Target: The Case of Aminoglycosides. *ChemBioChem* **2003**, *4* (10), 1018–1023. https://doi.org/10.1002/cbic.200300684.

(79) Swaney, S. The Oxazolidinone Linezolid Inhibits Initiation of Protein Synthesis in Bacteria. *Antimicrob. Agents Chemother.*

(80) Ippolito, J. A.; Kanyo, Z. F.; Wang, D.; Franceschi, F. J.; Moore, P. B.; Steitz, T. A.; Duffy, E. M. Crystal Structure of the Oxazolidinone Antibiotic Linezolid Bound to the 50S Ribosomal Subunit. *J. Med. Chem.* **2008**, *51* (12), 3353–3356. https://doi.org/10.1021/jm800379d.

(81) Elucidation of the RNA Target of Linezolid by Using a Linezolid–neomycin B Heteroconjugate and Genomic SELEX. *Bioorg. Med. Chem.* **2007**, *15* (24), 7688–7695. https://doi.org/10.1016/j.bmc.2007.08.053.

(82) Cross-Linking in the Living Cell Locates the Site of Action of Oxazolidinone Antibiotics. *J. Biol. Chem.* **2003**, *278* (24), 21972–21979. https://doi.org/10.1074/jbc.M302109200.

(83) Lirui, G.; Matthew D., D. Recent Advances in Developing Small Molecules Targeting RNA. *ACS Chem. Biol.* **2012**, *7*, 73–86. https://doi.org/10.1021/cb200447r.

(84) Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principles for Targeting RNA with Drug-like Small Molecules. *Nat. Rev. Drug Discov.* **2018**. https://doi.org/10.1038/nrd.2018.93.

(85) Rizvi, N. F.; Smith, G. F. RNA as a Small Molecule Druggable Target. *Bioorg. Med. Chem. Lett.* **2017**. https://doi.org/10.1016/j.bmcl.2017.10.052.

(86) Ursu, A.; Childs-Disney, J. L.; Andrews, R. J.; O'Leary, C. A.; Meyer, S. M.; Angelbello, A. J.; Moss, W. N.; Disney, M. D. Design of Small Molecules Targeting RNA Structure from Sequence. *Chem. Soc. Rev.* **2020**, *49* (20), 7252–7270. https://doi.org/10.1039/D0CS00455C.

(87) Hewitt, W. M.; Calabrese, D. R.; Schneekloth, J. S. Evidence for Ligandable Sites in Structured RNA throughout the Protein Data Bank. *Bioorg. Med. Chem.* **2019**, *27* (11), 2253–2260. https://doi.org/10.1016/j.bmc.2019.04.010.

(88) Zamore, P. D.; Tuschl, T.; Sharp, P. A.; Bartel, D. P. RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of MRNA at 21 to 23 Nucleotide Intervals. *Cell* **2000**, *101* (1), 25–33. https://doi.org/10.1016/S0092-8674(00)80620-0.

(89) Miura, F.; Kawaguchi, N.; Yoshida, M.; Uematsu, C.; Kito, K.; Sakaki, Y.; Ito, T. Absolute Quantification of the Budding Yeast Transcriptome by Means of Competitive PCR between Genomic and Complementary DNAs. *BMC Genomics* **2008**, *9*, 574. https://doi.org/10.1186/1471-2164-9-574.

(90) Gregory, S. G.; Barlow, K. F.; McLay, K. E.; Kaul, R.; Swarbreck, D.; Dunham, A.; Scott, C. E.; Howe, K. L.; Woodfine, K.; Spencer, C. C. A.; Jones, M. C.; Gillson, C.; Searle, S.; Zhou, Y.; Kokocinski, F.; McDonald, L.; Evans, R.; Phillips, K.; Atkinson, A.; Cooper, R.; Jones, C.; Hall, R. E.; Andrews, T. D.; Lloyd, C.; Ainscough, R.; Almeida, J. P.; Ambrose, K. D.; Anderson, F.; Andrew, R. W.; Ashwell, R. I. S.; Aubin, K.; Babbage, A. K.; Bagguley, C. L.; Bailey, J.; Beasley, H.; Bethel, G.; Bird, C. P.; Bray-Allen, S.; Brown, J. Y.; Brown, A. J.; Buckley, D.; Burton, J.; Bye, J.; Carder, C.; Chapman, J. C.; Clark, S. Y.; Clarke, G.; Clee, C.; Cobley, V.; Collier, R. E.; Corby, N.; Coville, G. J.; Davies, J.; Deadman, R.; Dunn, M.; Earthrowl, M.; Ellington, A. G.; Errington, H.; Frankish, A.; Frankland, J.; French, L.; Garner, P.; Garnett, J.; Gay, L.; Ghori, M. R. J.; Gibson, R.; Gilby, L. M.; Gillett, W.; Glithero, R. J.; Grafham, D. V.; Griffiths, C.; Griffiths-Jones, S.; Grocock, R.; Hammond, S.; Harrison, E. S. I.; Hart, E.; Haugen, E.; Heath, P. D.; Holmes, S.; Holt, K.; Howden, P. J.; Hunt, A. R.; Hunt, S. E.; Hunter, G.; Isherwood, J.; James, R.; Johnson, C.; Johnson, D.; Joy, A.; Kay, M.; Kershaw, J. K.; Kibukawa, M.; Kimberley, A. M.; King, A.; Knights, A. J.; Lad, H.; Laird, G.; Lawlor, S.; Leongamornlert, D. A.; Lloyd, D. M.; Loveland, J.; Lovell, J.; Lush, M. J.; Lyne, R.; Martin, S.; Mashreghi-Mohammadi, M.; Matthews, L.; Matthews, N. S. W.; McLaren, S.; Milne, S.; Mistry, S.; Oore, M. J. F. M.; Nickerson, T.; O'Dell, C. N.; Oliver, K.; Palmeiri, A.; Palmer, S. A.; Parker, A.; Patel, D.; Pearce, A. V.; Peck, A. I.; Pelan, S.; Phelps, K.; Phillimore, B. J.; Plumb, R.; Rajan, J.; Raymond, C.; Rouse, G.; Saenphimmachak, C.; Sehra, H. K.; Sheridan, E.; Shownkeen, R.; Sims, S.; Skuce, C. D.; Smith, M.; Steward, C.; Subramanian, S.; Sycamore, N.; Tracey, A.; Tromans, A.; Helmond, Z. V.; Wall, M.; Wallis, J. M.; White, S.; Whitehead, S. L.; Wilkinson, J. E.; Willey, D. L.; Williams, H.; Wilming, L.; Wray, P. W.; Wu, Z.; Coulson, A.; Vaudin, M.; Sulston, J. E.; Durbin, R.; Hubbard, T.; Wooster, R.; Dunham, I.; Carter, N. P.; McVean, G.; Ross, M. T.; Harrow, J.; Olson, M. V.; Beck, S.; Rogers, J.; Bentley, D. R. The DNA Sequence and Biological Annotation of Human Chromosome 1. *Nature* **2006**, *441* (7091), 315–321. https://doi.org/10.1038/nature04727.

(91) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*, 7th ed.; W. H. Freeman and Company.

(92) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* **2000**, *289* (5481), 920–930. https://doi.org/10.1126/science.289.5481.920.

(93) Martick, M.; Scott, W. G. Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis. *Cell* **2006**, *126* (2), 309–320. https://doi.org/https://doi.org/10.1016/j.cell.2006.06.036.

(94) Laganà, A.; Veneziano, D.; Russo, F.; Pulvirenti, A.; Giugno, R.; Croce, C. M.; Ferro, A. Computational Design of Artificial RNA Molecules for Gene Regulation. In *RNA Bioinformatics*; Methods in Molecular Biology; Humana Press, New York, NY, 2015; pp 393–412. https://doi.org/10.1007/978-1-4939-2291-8_25.

(95) Lünse, C. E.; Schüller, A.; Mayer, G. The Promise of Riboswitches as Potential Antibacterial Drug Targets. *Int. J. Med. Microbiol.* **2014**, *304* (1), 79–92. https://doi.org/10.1016/j.ijmm.2013.09.002.

(96) Panchal, V.; Brenk, R. Riboswitches as Drug Targets for Antibiotics. *Antibiotics* **2021**, *10* (1), 45. https://doi.org/10.3390/antibiotics10010045.

(97) Deigan, K.; Ferré-D'Amaré, A. R. Riboswitches: Discovery of Drugs That Target Bacterial Gene-Regulatory RNAs. *Acc. Chem. Res.* **2011**. https://doi.org/10.1021/ar200039b.

(98) Blouin, S.; Mulhbacher, J.; Penedo, J. C.; Lafontaine, D. A. Riboswitches: Ancient and Promising Genetic Regulators. *ChemBioChem* **2009**, *10* (3), 400–416. https://doi.org/10.1002/cbic.200800593.

(99) Blount, K. F.; Breaker, R. R. Riboswitches as Antibacterial Drug Targets. *Nat. Biotechnol.* **2006**, *24* (12), 1558–1564. https://doi.org/10.1038/nbt1268.

(100) Husser, C.; Dentz, N.; Ryckelynck, M. Structure-Switching RNAs: From Gene Expression Regulation to Small Molecule Detection. *Small Struct.* **2021**, *2* (4), 2000132. https://doi.org/10.1002/sstr.202000132.

(101) Winkler, W.; Nahvi, A.; Breaker, R. R. Thiamine Derivatives Bind Messenger RNAs Directly to Regulate Bacterial Gene Expression | Nature. *Nature* **2002**, *419*, 952–956. https://doi.org/https://doi.org/10.1038/nature01145.

(102) McCown, P. J.; Corbino, K. A.; Stav, S.; Sherlock, M. E.; Breaker, R. R. Riboswitch Diversity and Distribution. *RNA N. Y. N* **2017**, *23* (7), 995–1011. https://doi.org/10.1261/rna.061234.117.

(103) Pavlova, N.; Kaloudas, D.; Penchovsky, R. Riboswitch Distribution, Structure, and Function in Bacteria. *Gene* **2019**, *708*, 38–48. https://doi.org///doi.org/10.1016/j.gene.2019.05.036.

(104) Winkler, W. C.; Breaker, R. R. Regulation of Bacterial Gene Expression by Riboswitches. *Annu. Rev. Microbiol.* **2005**, *59* (1), 487–517. https://doi.org/10.1146/annurev.micro.59.030804.121336.

(105) Rekand, I.; Brenk, R. Design of Riboswitch Ligands, an Emerging Target Class for Novel Antibiotics. *Future Med Chem* **2017**, *9* (14), 1649–1662. https://doi.org/10.4155/fmc-2017-0063.

(106) Pavlova, N.; Penchovsky, R. Genome-Wide Bioinformatics Analysis of FMN, SAM-I, GlmS, TPP, Lysine, Purine, Cobalamin, and SAH Riboswitches for Their Applications as Allosteric Antibacterial Drug Targets in Human Pathogenic Bacteria. *Expert Opin. Ther. Targets* **2019**, *23* (7), 631–643. https://doi.org/10.1080/14728222.2019.1618274.

(107) Woolley, D. W.; White, A. G. C. Selective Reversible Inhibition of Microbial Growth with Pyrithiamine. *J. Exp. Med.* **1943**, *78* (6), 489–497. https://doi.org/10.1084/jem.78.6.489.

(108) Crystal Structures of the Thi-Box Riboswitch Bound to Thiamine Pyrophosphate Analogs Reveal Adaptive RNA-Small Molecule Recognition. *Structure* **2006**, *14* (9), 1459–1468. https://doi.org/10.1016/j.str.2006.07.008.

(109) Cressina, E.; Chen, L.; Abell, C.; Leeper, F. J.; Smith, A. G. Fragment Screening against the Thiamine Pyrophosphate RiboswitchthiM. *Chem. Sci.* **2010**, *2* (1), 157–165. https://doi.org/10.1039/C0SC00406E.

(110) Warner, K. D.; Homan, P.; Weeks, K. M.; Smith, A. G.; Abell, C.; Ferré-D'Amaré, A. R. Validating Fragment-Based Drug Discovery for Biological RNAs: Lead Fragments Bind and Remodel the TPP Riboswitch Specifically. *Chem. Biol.* **2014**, *21* (5), 591–595. https://doi.org/10.1016/j.chembiol.2014.03.007.

(111) Batey, R. T.; Gilbert, S. D.; Montange, R. K. Structure of a Natural Guanine-Responsive Riboswitch Complexed with the Metabolite Hypoxanthine. *Nature* **2004**. https://doi.org/10.1038/nature03037.

(112) Kim, J. N.; Blount, K. F.; Puskarz, I.; Lim, J.; Link, K. H.; Breaker, R. R. Design and Antimicrobial Action of Purine Analogues That Bind Guanine Riboswitches. *ACS Chem. Biol.* **2009**, *4* (11), 915–927. https://doi.org/10.1021/cb900146k.

(113) Mulhbacher, J.; Brouillette, E.; Allard, M.; Fortier, L.-C.; Malouin, F.; Lafontaine, D. A. Novel Riboswitch Ligand Analogs as Selective Inhibitors of Guanine-Related Metabolic Pathways. *PLOS Pathog.* **2010**, *6* (4), e1000865. https://doi.org/10.1371/journal.ppat.1000865.

(114) Shiota, T.; Mauron, J.; Folk, J. E. Inhibition of Lysine and Lysine Peptide Utilization in Bacteria by Peptides of S-(β-Aminoethyl)-Cysteine. *Biochim. Biophys. Acta* **1961**, *53* (2), 360–365. https://doi.org/10.1016/0006-3002(61)90447-4.

(115) McCord, T. J.; Ravel, J. M.; Skinner, C. G.; Shive, W. DL-4-Oxalysine, an Inhibitory Analog of Lysine. *J. Am. Chem. Soc.* **1957**, *79* (21), 5693–5696. https://doi.org/10.1021/ja01578a029.

(116) Blount, K. F.; Wang, J. X.; Lim, J.; Sudarsan, N.; Breaker, R. R. Antibacterial Lysine Analogs That Target Lysine Riboswitches. *Nat. Chem. Biol.* **2007**, *3* (1), 44–49. https://doi.org/https://doi.org/10.1038/nchembio842.

(117) Sudarsan, N.; Wickiser, J. K.; Nakamura, S.; Ebert, M. S.; Breaker, R. R. An MRNA Structure in Bacteria That Controls Gene Expression by Binding Lysine. *Genes Dev.* **2003**, *17* (21), 2688–2697. https://doi.org/10.1101/gad.1140003.

(118) Gelfand, M. S.; Mironov, A. A.; Jomantas, J.; Kozlov, Y. I.; Perumov, D. A. A Conserved RNA Structure Element Involved in the Regulation of Bacterial Riboflavin Synthesis Genes. *Trends Genet.* **1999**, *15* (11), 439–442. https://doi.org/10.1016/S0168-9525(99)01856-9.

(119) Winkler, W. C.; Cohen-Chalamish, S.; Breaker, R. R. An MRNA Structure That Controls Gene Expression by Binding FMN. *Proc. Natl. Acad. Sci.* **2002**, *99* (25), 15908–15913. https://doi.org/https://doi.org/10.1073/pnas.212628899.

(120) Mironov, A. S.; Gusarov, I.; Rafikov, R.; Lopez, L. E.; Shatalin, K.; Kreneva, R. A.; Perumov, D. A.; Nudler, E. Sensing Small Molecules by Nascent RNA: A Mechanism to Control Transcription in Bacteria. *Cell* **2002**, *111* (5), 747–756. https://doi.org/10.1016/S0092-8674(02)01134-0.

(121) Serganov, A.; Nudler, E. A Decade of Riboswitches. *Cell* **2013**, *152* (1), 17–24. https://doi.org/10.1016/j.cell.2012.12.024.

(122) Serganov, A.; Huang, L.; Patel, D. J. Coenzyme Recognition and Gene Regulation by a Flavin Mononucleotide Riboswitch. *Nature* **2009**, *458* (7235), 233–237. https://doi.org/10.1038/nature07642.

(123) Mack Matthias; van Loon Adolphus P. G. M.; Hohmann Hans-Peter. Regulation of Riboflavin Biosynthesis InBacillus Subtilis Is Affected by the Activity of the

Flavokinase/Flavin Adenine Dinucleotide Synthetase Encoded ByribC. *J. Bacteriol.* **1998**, *180* (4), 950–955. https://doi.org/10.1128/JB.180.4.950-955.1998.

(124) Fischer, M.; Bacher, A. Biosynthesis of Flavocoenzymes. *Nat. Prod. Rep.* **2005**, *22* (3), 324–350. https://doi.org/10.1039/B210142B.

(125) Gutiérrez-Preciado, A.; Torres, A. G.; Merino, E.; Bonomi, H. R.; Goldbaum, F. A.; García-Angulo, V. A. Extensive Identification of Bacterial Riboflavin Transporters and Their Distribution across Bacterial Species. *PLOS ONE* **2015**, *10* (5), e0126124. https://doi.org/10.1371/journal.pone.0126124.

(126) Lee, E. R.; Blount, K. F.; Breaker, R. R. Roseoflavin Is a Natural Antibacterial Compound That Binds to FMN Riboswitches and Regulates Gene Expression. *RNA Biol.* **2009**, *6* (2), 187–194. https://doi.org/10.4161/rna.6.2.7727.

(127) Mack, M.; Grill, S. Riboflavin Analogs and Inhibitors of Riboflavin Biosynthesis. *Appl. Microbiol. Biotechnol.* **2006**, *71* (3), 265–275. https://doi.org/10.1007/s00253-006-0421-7.

(128) Ott, E.; Stolz, J.; Mack, M. The RFN Riboswitch of Bacillus Subtilis Is a Target for the Antibiotic Roseoflavin Produced by Streptomyces Davawensis. *RNA Biol.* **2009**, *6* (3), 276–280. https://doi.org/10.4161/rna.6.3.8342.

(129) Kenneth F., B.; Cynthia, M.; Mark, P.; David, O.; Tim, O.; Paul, A.; Cheryl, Q.; R. Alan, C.; Toni J., P.; Heinrich J., S.; Catherine A., S.; Daniel P., W.; Peter G., W.; Ronald R., B. Novel Riboswitch-Binding Flavin Analog That Protects Mice against Clostridium Difficile Infection without Inhibiting Cecal Flora. *Antimicrob. Agents Chemother.* **2015**. https://doi.org/10.1128/AAC.01282-15.

(130) Vicens, Q.; Mondragón, E.; Batey, R. T. Molecular Sensing by the Aptamer Domain of the FMN Riboswitch: A General Model for Ligand Binding by Conformational Selection. *Nucleic Acids Res.* **2011**, *39* (19), 8586–8598. https://doi.org/10.1093/nar/gkr565.

(131) Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A.; Murgolo, N.; Barbieri, C. M.; Mann, P. A.; Carr, D.; Xia, E.; Zuck, P.; Riley, D.; Painter, R. E.; Walker, S. S.; Sherborne, B.; de Jesus, R.; Pan, W.; Plotkin, M. A.; Wu, J.; Rindgen, D.; Cummings, J.; Garlisi, C. G.; Zhang, R.; Sheth, P. R.; Gill, C. J.; Tang, H.; Roemer, T. Selective Small-Molecule Inhibition of an RNA Structural Element. *Nature* **2015**, *526* (7575), 672–677. https://doi.org/10.1038/nature15542.

(132) Howe, J. A.; Xiao, L.; Fischmann, T. O.; Wang, H.; Tang, H.; Villafania, A.; Zhang, R.; Barbieri, C. M.; Roemer, T. Atomic Resolution Mechanistic Studies of Ribocil: A Highly Selective Unnatural Ligand Mimic of the E. Coli FMN Riboswitch. *RNA Biol.* **2016**, *13* (10), 946–954. https://doi.org/10.1080/15476286.2016.1216304.

(133) Wehler, T. Structure-Based Design for the Discovery of Novel Ligands Targeting the FMN Riboswitch. Ph.D. Thesis, the Johannes Gutenberg University of Mainz, Department of Chemistry, Pharmacy and Geoscience, 2017.

(134) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; KDD '16; ACM: New York, NY, USA, 2016; pp 785–794. https://doi.org/10.1145/2939672.2939785.

(135) Rodríguez-Pérez, R.; Bajorath, J. Interpretation of Machine Learning Models Using Shapley Values: Application to Compound Potency and Multi-Target Activity Predictions. *J. Comput. Aided Mol. Des.* **2020**, *34* (10), 1013–1026. https://doi.org/10.1007/s10822-020-00314-0.

(136) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand Efficiency: A Useful Metric for Lead Selection. *Drug Discov. Today* **2004**, *9* (10), 430–431. https://doi.org/https://doi.org/10.1016/S1359-6446(04)03069-7.

(137) Flinders, J.; DeFina, S. C.; Brackett, D. M.; Baugh, C.; Wilson, C.; Dieckmann, T. Recognition of Planar and Nonplanar Ligands in the Malachite Green–RNA Aptamer Complex. *ChemBioChem* **2004**, *5* (1), 62–72. https://doi.org/10.1002/cbic.200300701.

(138) A Small-Molecule Probe Induces a Conformation in HIV TAR RNA Capable of Binding Drug-Like Fragments. *J. Mol. Biol.* **2011**, *410* (5), 984–996. https://doi.org/10.1016/j.jmb.2011.03.039.

(139) Sztuba-Solinska, J.; Shenoy, S. R.; Gareiss, P.; Krumpe, L. R. H.; Le Grice, S. F. J.; O'Keefe, B. R.; Schneekloth, J. S. Identification of Biologically Active, HIV TAR RNA-Binding Small Molecules Using Small Molecule Microarrays. *J. Am. Chem. Soc.* **2014**, *136* (23), 8402–8410. https://doi.org/10.1021/ja502754f.

(140) Yang K; Chang Jy; Cui Z; Li X; Meng R; Duan L; Thongchol J; Jakana J; Huwe Cm; Sacchettini Jc; Zhang J. Structural Insights into Species-Specific Features of the Ribosome from the Human Pathogen Mycobacterium Tuberculosis. *Nucleic Acids Res.* **2017**. https://doi.org/10.1093/nar/gkx785.

(141) Lin, A. H.; Murray, R. W.; Vidmar, T. J.; Marotti, K. R. The Oxazolidinone Eperezolid Binds to the 50S Ribosomal Subunit and Competes with Binding of Chloramphenicol and Lincomycin. *Antimicrob. Agents Chemother.* **1997**, *41* (10), 2127–2131. https://doi.org/https://doi.org/10.1128/AAC.41.10.2127.

(142) Belousoff, M. J.; Venugopal, H.; Wright, A.; Seoner, S.; Stuart, I.; Stubenrauch, C.; Bamert, R. S.; Lupton, D. W.; Lithgow, T. CryoEM-Guided Development of Antibiotics for Drug-Resistant Bacteria. *ChemMedChem* **2019**, *14* (5), 527–531. https://doi.org/10.1002/cmdc.201900042.

(143) Belousoff Mj; Wright A; Seoner S; Stuart I; Stubenrauch C; Bamert Rs; Lupton Dw; Lithgow T. cryoEM-Guided Development of Antibiotics for Drug-Resistant Bacteria https://pubmed.ncbi.nlm.nih.gov/30667174/. https://doi.org/10.1002/cmdc.201900042.

(144) Loubresse, N. G. de; Prokhorova, I.; Holtkamp, W.; Rodnina, M. V.; Yusupova, G.; Yusupov, M. Structural Basis for the Inhibition of the Eukaryotic Ribosome. *Nature* **2014**, *513* (7519), 517–522. https://doi.org/10.1038/nature13737.

(145) Blaha, G.; Gürel, G.; Schroeder, S. J.; Moore, P. B.; Steitz, T. A. Mutations Outside the Anisomycin-Binding Site Can Make Ribosomes Drug-Resistant. *J. Mol. Biol.* **2008**, *379* (3), 505–519. https://doi.org/10.1016/j.jmb.2008.03.075.

(146) Lee, M.-K.; Bottini, A.; Kim, M.; Bardaro, M. F.; Zhang, Z.; Pellecchia, M.; Choi, B.-S.; Varani, G. A Novel Small-Molecule Binds to the Influenza A Virus RNA Promoter and Inhibits Viral Replication. *Chem. Commun.* **2013**, *50* (3), 368–370. https://doi.org/10.1039/C3CC46973E.

(147) Warner, K. D.; Chen, M. C.; Song, W.; Strack, R. L.; Thorn, A.; Jaffrey, S. R.; Ferré-D'Amaré, A. R. Structural Basis for Activity of Highly Efficient RNA Mimics of

Green Fluorescent Protein. *Nat. Struct. Mol. Biol.* **2014**, *21* (8), 658–663. https://doi.org/10.1038/nsmb.2865.

(148) Edwards, T. E.; Ferré-D'Amaré, A. R. Crystal Structures of the Thi-Box Riboswitch Bound to Thiamine Pyrophosphate Analogs Reveal Adaptive RNA-Small Molecule Recognition. *Structure* **2006**, *14* (9), 1459–1468. https://doi.org/10.1016/j.str.2006.07.008.

(149) Wehrhan, L.; Hillisch, A.; Mundt, S.; Tersteegen, A.; Meier, K. Druggability Assessment for Selected Serine Proteases in a Pharmaceutical Industry Setting. *ChemMedChem* **2020**, *15* (21), 2010–2018. https://doi.org/https://doi.org/10.1002/cmdc.202000425.

(150) Lindemann, M.; Moldovan, R.-P.; Hinz, S.; Deuther-Conrad, W.; Gündel, D.; Dukic-Stefanovic, S.; Toussaint, M.; Teodoro, R.; Juhl, C.; Steinbach, J.; Brust, P.; Müller, C. E.; Wenzel, B. Development of a Radiofluorinated Adenosine A2B Receptor Antagonist as Potential Ligand for PET Imaging. *Int. J. Mol. Sci.* **2020**, *21* (9), 3197. https://doi.org/10.3390/ijms21093197.

(151) Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the Chemical Beauty of Drugs. *Nat. Chem.* **2012**, *4* (2), 90–98. https://doi.org/10.1038/nchem.1243.

(152) Burova, V. Chartering Chemical Space of Riboswitch Ligands Towards Future Antibiotics. Master Thesis in Chemistry, University of Bergen, Department of Chemistry, 2021.

(153) Vintermyr, H. K. Synthesis of Riboswitch Ligands. Master Thesis in Pharmacy, University of Bergen, Department of Chemistry, 2020.

(154) Årdal, A. O. Synthesis of Riboswitch Ligands. Master Thesis in Pharmacy, University of Bergen, Department of Chemistry, 2021.

(155) Muhammed, Z.; Haug, B.-E. Unpublished Results, Unversity of Bergen, Department of Chemistry.

(156) Moneam, M. I. A.; Geies, A. A.; El-Naggar, G. M.; Mousa, S. M. Synthesis of New Furo[2,3-d]Pyrimidines and Pyrimido[4′,5′:4,5]Furo[2,3-d]Pyrimidines. *J. Chin. Chem. Soc.* **2004**, *51* (6), 1357–1366. https://doi.org/10.1002/jccs.200400198.

(157) El-Dean, K.; M, A. Synthesis of Some Pyrimidothienopyrimidine Derivatives. *Monatshefte Für Chem. Chem. Mon.* **1998**, *129* (5), 523–533. https://doi.org/10.1007/PL00000109.

(158) Schaefer, J. P.; Bloomfield, J. J. The Dieckmann Condensation (Including the Thorpe-Ziegler Condensation). In *Organic Reactions*; John Wiley & Sons, Ltd, 2011; pp 1–203. https://doi.org/10.1002/0471264180.or015.01.

(159) Synthesis of Amino Derivatives of Five-Membered Heterocycles by Thorpe-Ziegler Cyclization. **1998**, *72*, 79–125. https://doi.org/10.1016/S0065-2725(08)60315-3.

(160) Sleebs, B. E.; Levit, A.; Street, I. P.; Falk, H.; Hammonds, T.; Wong, A. C.; Charles, M. D.; Olson, M. F.; Baell, J. B. Identification of 3-Aminothieno[2,3-b]Pyridine-2-Carboxamides and 4-Aminobenzothieno[3,2-d]Pyrimidines as LIMK1 Inhibitors. *MedChemComm* **2011**, *2* (10), 977–981. https://doi.org/10.1039/C1MD00137J.

(161) Keshk, R. M. Design and Synthesis of New Series of 3-Cyanopyridine and Pyrazolopyridine Derivatives. *J. Heterocycl. Chem.* **2020**, *57* (9), 3384–3393. https://doi.org/10.1002/jhet.4058.

(162) A. Back, R.; C. Boden, J. High-Temperature Photolysis and the Pyrolysis of Formamide Vapour, and the Thermal Decomposition of the Carbamyl Radical. *Trans. Faraday Soc.* **1971**, *67* (0), 88–96. https://doi.org/10.1039/TF9716700088.

(163) Cataldo, F.; Patanè, G.; Compagnini, G. Synthesis of HCN Polymer from Thermal Decomposition of Formamide. *J. Macromol. Sci. Part A* **2009**, *46* (11), 1039–1048. https://doi.org/10.1080/10601320903245342.

(164) Anastas, P.; Eghbali, N. Green Chemistry: Principles and Practice. *Chem. Soc. Rev.* **2010**, *39* (1), 301–312. https://doi.org/10.1039/B918763B.

(165) Evans, D. Preparation of Pyrazolopyridines and Pyrrolopyridines as Semicarbazide-Sensitive Amine Oxidase Inhibitors. WO2011113798.

(166) Sizova, O. S.; Modnikova, G. A.; Glushkov, R. G.; Solov'eva, N. P.; Ryabokon', N. A.; Chernov, V. A.; Okinshevich, O. V.; Pershin, G. N. Synthesis and Biological Activity of 4,7-Substituted Pyrrolo[3,2-d]Pyrimidines. *Pharm. Chem. J.* **1984**, *18* (8), 567–571. https://doi.org/10.1007/BF00779278.

(167) Baeschlin, D. K.; Clark, D. E.; Dunsdon, S. J.; Fenton, G.; Fillmore, A.; Harris, N. V.; Higgs, C.; Hurley, C. A. Condensed Heterocyclic Compounds Useful as DPP-IV Inhibitors and Their Preparation. WO2007071738A1.

(168) Kataoka, K.; Kosugi, T.; Ishii, T.; Takeuchi, T. Preparation of Pyrrolopyrimidine Derivatives as GSK-3 Inhibitors. WO2003070729.

(169) Chen, N.; Lu, Y.; Gadamasetti, K.; Hurt, C. R.; Norman, M. H.; Fotsch, C. A Short, Facile Synthesis of 5-Substituted 3-Amino-1H-Pyrrole-2-Carboxylates. *J. Org. Chem.* **2000**, *65* (8), 2603–2605. https://doi.org/10.1021/jo9917902.

(170) Jung, Y.-H.; Yu, J.; Wen, Z.; Salmaso, V.; Karcz, T. P.; Phung, N. B.; Chen, Z.; Duca, S.; Bennett, J. M.; Dudas, S.; Salvemini, D.; Gao, Z.-G.; Cook, D. N.; Jacobson, K. A. Exploration of Alternative Scaffolds for P2Y14 Receptor Antagonists Containing a Biaryl Core. *J. Med. Chem.* **2020**, *63* (17), 9563–9589. https://doi.org/10.1021/acs.jmedchem.0c00745.

(171) Banfi, D.; Patiny, L. Www.Nmrdb.Org: Resurrecting and Processing NMR Spectra On-Line. *CHIMIA* **2008**, *62* (4), 280. https://doi.org/10.2533/chimia.2008.280.

(172) Steinbeck, C.; Krause, S.; Kuhn, S. NMRShiftDBConstructing a Free Chemical Information System with Open-Source Components. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1733–1739. https://doi.org/10.1021/ci0341363.

(173) Aprà, E.; Bylaska, E. J.; de Jong, W. A.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; van Dam, H. J. J.; Alexeev, Y.; Anchell, J.; Anisimov, V.; Aquino, F. W.; Atta-Fynn, R.; Autschbach, J.; Bauman, N. P.; Becca, J. C.; Bernholdt, D. E.; Bhaskaran-Nair, K.; Bogatko, S.; Borowski, P.; Boschen, J.; Brabec, J.; Bruner, A.; Cauët, E.; Chen, Y.; Chuev, G. N.; Cramer, C. J.; Daily, J.; Deegan, M. J. O.; Dunning, T. H.; Dupuis, M.; Dyall, K. G.; Fann, G. I.; Fischer, S. A.; Fonari, A.; Früchtl, H.; Gagliardi, L.; Garza, J.; Gawande, N.; Ghosh, S.; Glaesemann, K.; Götz, A. W.; Hammond, J.; Helms, V.; Hermes, E. D.; Hirao, K.; Hirata, S.; Jacquelin, M.; Jensen, L.; Johnson, B. G.; Jónsson, H.; Kendall, R. A.; Klemm, M.; Kobayashi, R.; Konkov, V.; Krishnamoorthy, S.; Krishnan, M.; Lin, Z.; Lins, R. D.; Littlefield, R. J.; Logsdail, A. J.; Lopata, K.; Ma, W.; Marenich, A. V.; Martin del Campo, J.; Mejia-Rodriguez, D.; Moore, J. E.; Mullin, J. M.; Nakajima, T.; Nascimento, D. R.; Nichols, J. A.; Nichols, P. J.; Nieplocha, J.; Otero-de-la-Roza, A.; Palmer, B.; Panyala, A.;

Pirojsirikul, T.; Peng, B.; Peverati, R.; Pittner, J.; Pollack, L.; Richard, R. M.; Sadayappan, P.; Schatz, G. C.; Shelton, W. A.; Silverstein, D. W.; Smith, D. M. A.; Soares, T. A.; Song, D.; Swart, M.; Taylor, H. L.; Thomas, G. S.; Tipparaju, V.; Truhlar, D. G.; Tsemekhman, K.; Van Voorhis, T.; Vázquez-Mayagoitia, Á.; Verma, P.; Villa, O.; Vishnu, A.; Vogiatzis, K. D.; Wang, D.; Weare, J. H.; Williamson, M. J.; Windus, T. L.; Woliński, K.; Wong, A. T.; Wu, Q.; Yang, C.; Yu, Q.; Zacharias, M.; Zhang, Z.; Zhao, Y.; Harrison, R. J. NWChem: Past, Present, and Future. *J. Chem. Phys.* **2020**, *152* (18), 184102. https://doi.org/10.1063/5.0004997.

(174) Jacquemard, U.; Beneteau, V.; Lefoix, M.; Routier, S.; Merour, J.-Y.; Coudert, G. Mild and Selective Deprotection of Carbamates with Bu4NF. *Tetrahedron* **2004**, *60* (44), 10039–10047. https://doi.org/10.1016/j.tet.2004.07.071.

(175) McIsaac, J. E.; Ball, R. E.; Behrman, E. J. Mechanism of the Base-Catalyzed Conversion of Nitriles to Amides by Hydrogen Peroxide. *J. Org. Chem.* **1971**, *36* (20), 3048–3050. https://doi.org/10.1021/jo00819a034.

(176) Clayton, R.; Davis, M. L.; Li, W.; Fraser, W.; Ramsden, C. A. Preparation of Pyridine-Stretched 2'-Deoxyhypoxanthosine Phosphoramidite. *ARKIVOC* **2017**, *2017* (3), 87–104. https://doi.org/10.24820/ark.5550190.p010.075.

(177) Matos, K.; Soderquist, J. A. Alkylboranes in the Suzuki−Miyaura Coupling: Stereochemical and Mechanistic Studies. *J. Org. Chem.* **1998**, *63* (3), 461–470. https://doi.org/10.1021/jo971681s.

(178) Patching, S. G. Surface Plasmon Resonance Spectroscopy for Characterisation of Membrane Protein–ligand Interactions and Its Potential for Drug Discovery. *Biochim. Biophys. Acta BBA - Biomembr.* **2014**, *1838* (1), 43–55. https://doi.org/10.1016/j.bbamem.2013.04.028.

(179) Rekand, I. H.; Brenk, R. DrugPred_RNA—A Tool for Structure-Based Druggability Predictions for RNA Binding Sites. *J. Chem. Inf. Model.* **2021**, *61* (8), 4068–4081. https://doi.org/10.1021/acs.jcim.1c00155.

(180) A Practical Palladium Catalyzed Dehalogenation of Aryl Halides and α-Haloketones. *Tetrahedron* **2007**, *63* (20), 4266–4270. https://doi.org/10.1016/j.tet.2007.03.061.

(181) Churchill, G. H.; Raw, S. A.; Powell, L. Improved Synthesis of Substituted Pyrido[2,3-d]Pyrimidinediones. *Tetrahedron Lett.* **2011**, *52* (28), 3657–3661. https://doi.org/10.1016/j.tetlet.2011.05.021.

(182) Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. L. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25* (11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163.

(183) Landrum, G. RDKit: Open-source cheminformatics http://www.rdkit.org.

(184) Sarkar, A.; Brenk, R. To Hit or Not to Hit, That Is the Question - Genome-Wide Structure-Based Druggability Predictions for Pseudomonas Aeruginosa Proteins. *PloS One* **2015**, *10* (9), e0137279. https://doi.org/10.1371/journal.pone.0137279.

(185) Lorber, D. M.; Shoichet, B. K. Flexible Ligand Docking Using Conformational Ensembles. *Protein Sci.* **1998**, *7* (4), 938–50. https://doi.org/https://doi.org/10.1002/pro.5560070411.

(186) Mitternacht, S. FreeSASA: An Open Source C Library for Solvent Accessible Surface Area Calculations. *F1000Research* **2016**, *5*. https://doi.org/10.12688/f1000research.7931.1.

(187) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. The Packing Density in Proteins: Standard Radii and Volumes11Edited by J. M. Thornton. *J. Mol. Biol.* **1999**, *290* (1), 253–266. https://doi.org/10.1006/jmbi.1999.2829.

(188) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.

(189) Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*; Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc., 2017; pp 4765–4774.

(190) Liu, Y.; Just, A. *SHAPforxgboost: SHAP Plots for "XGBoost";* manual; 2019.

(191) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15* (3), 1863–1874. https://doi.org/10.1021/acs.jctc.8b01026.

(192) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50* (4), 572–584. https://doi.org/10.1021/ci100031x.

(193) Mysinger, M. M.; Shoichet, B. K. Rapid Context-Dependent Ligand Desolvation in Molecular Docking. *J. Chem. Inf. Model.* **2010**, *50* (9), 1561–1573. https://doi.org/10.1021/ci100214a.

(194) Peter, D.; Reyes, F. E.; Robinson, D. A.; Hammond, C. A.; Lilley, D. M.; Batey, R. T.; Brenk, R. Novel Ligands for a Purine Riboswitch Discovered by RNA-Ligand Docking. *Chem. Biol.* **2011**, *18* (3), 324–335. https://doi.org/10.1016/j.chembiol.2010.12.020.

(195) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (5), 868–873. https://doi.org/10.1021/ci990307l.

(196) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. Computational Prediction of 1H and 13C Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry. *Chem. Rev.* **2012**, *112* (3), 1839–1862. https://doi.org/10.1021/cr200106v.

(197) CHESHIRE Chemical Shift Repository http://cheshirenmr.info/Recommendations.htm#tantilloccs (accessed 2022 -03 -25).

Errata for
“Structure-based Druggability Predictions of RNA Binding Sites and Design of RNA Ligands

Illimar Rekand

Thesis for the degree philosophiae doctor (PhD)
at the University of Bergen

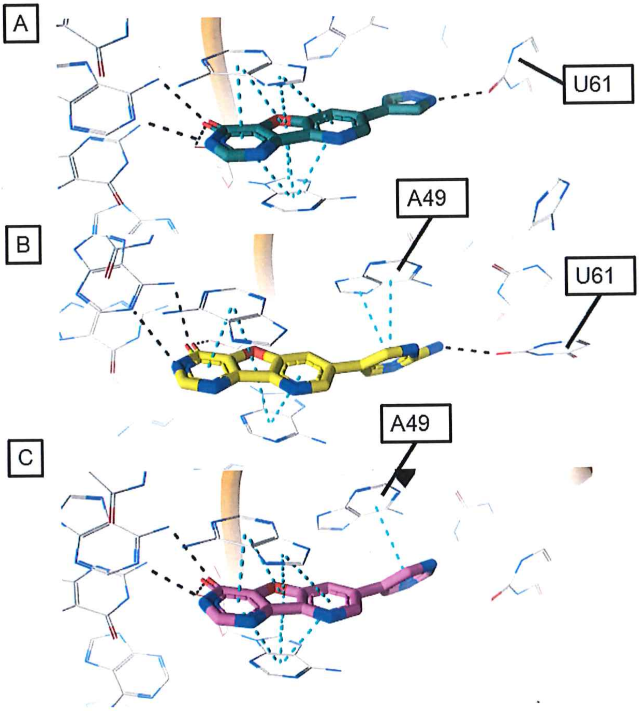
18.08.22
(date and sign. of candidate)

Martha Enger 19.08.22
(date and sign. of faculty)

Errata

Page 69 Correction in the text: "In the modelled binding mode, **RSL-0029** formed a hydrogen bond with the carbonyl oxygen in U61 (**Figure 22A**), while **RSL-0026** forms both a hydrogen bond with the same residue (**Figure 22B**). Both **RSL-0026** and **RSL-0027** (**Figure 22C**) form a stacking π/π-interaction with A49. These compounds were selected for further evaluation.

Page 69 Incorrect figure



Page 76 Remove: Molecular minimization suggested that **RSL-0034** formed an edge-face interaction with G62 (Error! Reference source not found.**CD**~~, similar to what has previously been seen with **RSL-0029**, **RSL-0026**, (Error! Reference source not found.**A, C**), and **RSL-0007** (Error! Reference source not found.**C**),~~ and also a hydrogen bond with the endocyclic ribose oxygen of A85, not observed previously.

Graphic design: Communication Division, UIB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230850268 (print)
9788230866047 (PDF)