

Combining Primary Specificity Screenings for  
Drug Discovery Targeting T-box Antiterminator RNA

---

A Thesis

Presented to

The Honors Tutorial College

Ohio University

---

In Partial Fulfillment

of the Requirements for Graduation

from the Honors Tutorial College

with the degree of

Bachelor of Science in Chemistry

---

by

Mason T. Myers

April 2021

This Thesis titled  
Combining Primary Specificity Screenings for Drug Discovery Targeting T-box  
Antiterminator RNA

By  
Mason T. Myers

Has been approved for  
the Department of Chemistry and Biochemistry  
and the Honors Tutorial College

---

Dr. Jennifer V. Hines

Thesis Adviser  
Professor, Chemistry and Biochemistry

---

Dr. Lauren E. H. McMills  
Director of Studies, Chemistry and Biochemistry

---

Dr. Donal Skinner  
Dean, Honors Tutorial College

## Abstract

Myers, Mason T., April 2021, Chemistry

Combining Primary Specificity Screenings for Drug Discovery Targeting T-box

Antiterminator RNA

Director of Thesis: Jennifer V. Hines

As the threat of antibiotic resistant infections and outbreaks looms, there has been a reinvigorated interest in identifying new therapeutics to target alternative targets in species primed for developing resistance. One such target is the antiterminator sequence of the T-box riboswitch, an important regulatory motif that acts as an ‘on’ switch for important protein synthesis genes in Gram-positive bacteria. The antiterminator is kinetically favored in transcription of T-box regulated genes, but is thermodynamically disfavored to its counterpart terminator sequence, which contains many of the same nucleotides and prevents gene expression through transcription termination or enveloping the Shine-Dalgarno sequence, preventing translation. The antiterminator is stabilized through interaction with the acceptor end of uncharged tRNA, and as such is responsive to the cellular concentration ratio of charged and uncharged tRNA. As a thermodynamically unstable and highly conserved regulatory element, the T-box antiterminator has been the focus of drug-design efforts to create ligands that would preclude or destabilize tRNA binding to the antiterminator, disrupting protein biogenesis ultimately leading to cell death.

In an effort to devise a new primary, high-moderate throughput compound screening to find small molecules which bind to the antiterminator mechanism of the T-box riboswitch with high specificity, this thesis investigates a hybrid assay combining

computational and experimental techniques. Computational docking of libraries of compounds using a receptor grid developed from the antiterminator NMR solution structure (PDB: 1N53) is used to identify a selection of compounds with favorable chemical features which bind to the antiterminator with high selectivity and strong bonding values. These compounds can then be tested in a single temperature fluorescence assay against three similar, but structurally disparate models based on the T-box antiterminator to identify ligand affinity and binding specificity, important aspects of drug discovery research. Agreement between computational and experimental techniques will lead to the identification of common structures or trends in the molecules tested which effectively bind and modulate the antiterminator structure, enhancing the foundational knowledge required for pharmacophore development of a T-box antiterminator inhibitor.

Two compound libraries, the MedChemExpress FDA-approved plus library and the ZINC database natural metabolite subset, were tested using this combined assay approach. In addition, a selection of laboratory compounds known to bind RNA were also tested in an inverse sequence of the assay, completing fluorescence screenings followed by computational assays. Of the candidate molecules identified in computational screenings, four compounds were tested in the experimental assays, two from each library. One of these compounds, amodiaquine, was found bind to the T-box antiterminator with structure -dependent specificity, and the compounds screened from the laboratory 'training' set all had structure-dependent specificity. The results of this project indicate that while specificity is not well determined by computational screening of a compound library, annotation of compound interaction by receptor regions improves

prediction of relative ligand binding strengths. In addition, the quinoline ring system appears to be an intriguing moiety for RNA drug design, appearing in multiple compounds that bind the antiterminator to affect its structure in a model-dependent manner. In sum, the results of this assay support the combination of computational and experimental assays to better predict RNA-small molecule binding in drug-discovery efforts targeting the T-box riboswitch.

## Acknowledgements

I would like to extend my foremost thanks and gratitude to my thesis advisor and research mentor, Dr. Jennifer Hines, who has provided me guidance and opportunity through the entirety of my undergraduate studies. Thank you, Dr. Hines, for allowing me to join your lab so early, encouraging me to tackle new projects, present my results, and giving me both the tutelage and space to let me learn and grow as a researcher and scientist, through accomplishments and mistakes alike. Thank you also for the many hours spent writing recommendations and revising applications for internships, grants, graduate school, and this thesis too. I am grateful for all you have done through your mentorship to help me develop my abilities as a scientist. I am grateful for my lab mates, who always brought teamwork, enthusiasm, and an enjoyable atmosphere to our lab. I want to thank Ismail and Ali for their wisdom, patience, and willingness to show me the ropes in RNA and drug development research. You both were a constant source of advice and encouragement in my coursework and research, and I can't have asked for a better pair of group members to have learned from and collaborated with these past four years. I must also thank Emily, Brianna, Shifra, and Eric for their efforts, perspectives, and contributions that have helped our research to expand and develop and made the lab a great place to work in. I want to thank the Chemistry and Biochemistry faculty for the foundational knowledge and skills they have taught me through coursework and tutorials. In particular, thank you, Dr. Lauren McMills and Dr. Mark McMills, for all of your advice, encouragement, and small chats in your offices that have helped me progress as a chemist. Lastly, I must thank my parents, Laura and Todd Myers, for their support through my childhood and college years.

## Table of Contents

Abstract.....	3
Acknowledgements .....	6
Table of Contents .....	7
List of Table and Figures.....	11
1. Introduction .....	14
1.1 The T-box Riboswitch .....	15
1.2 Structure and mechanism of the T-box Riboswitch .....	17
1.3 Ligand identification and binding to the T-box riboswitch .....	21
1.4 RNA drug discovery and design .....	23
2. Screening Design.....	33
2.1 Computation setup and evaluation .....	33
2.1.1 User Input .....	33
2.1.2 Computation .....	34
2.1.3 Evaluation.....	36
2.2 Experimental design and analysis .....	38
2.2.1 Fluorescence Experimentation – Control Screen .....	38
2.2.2 Analysis – Control Screen.....	38
2.2.3 Fluorescence Experimentation – Reproducibility Screen .....	39
2.2.4 Analysis – Reproducibility Screen .....	39

2.2.5	Fluorescence Experimentation – Specificity Screen .....	40
2.2.6	Analysis – Specificity Screen .....	41
3.	Results – FDA-Approved Library .....	43
3.1	Computational preparation and docking yields potential antiterminator ligands in an FDA drug-repositioning library with high interaction scores and positional consistency .....	43
3.2	Experimental screenings of FDA compounds indicate binding specificity in amodiaquine interactions with the antiterminator .....	49
4.	Results – ZINC Natural Metabolite Library .....	53
4.1	Computational preparation and docking of the ZINC library determined a subset of soluble, high affinity and positionally consistent compounds .....	53
4.2	Experimental screenings of ZINC compounds indicated nonspecific interactions with the antiterminator .....	58
5.	Results – Lab Compound Training Set .....	61
5.1	Experimental screenings of polyamines displayed antiterminator model-specific fluorescence modulation.....	62
5.2	Experimental screenings of polycyclic heterocycles identified antiterminator model-specific fluorescence in all tested compounds .....	64
5.3	Training set compounds were computationally docked to AM1A.....	66
6.	Results comparison of computational studies to fluorescence results.....	68
6.1	Development and docking for a computational AM <sub>Control</sub> model .....	68

6.2 Comparison of computational and Emodel values reveals trends in predictive value of computational docking by region.....	69
7. Conclusion.....	73
8. Methods.....	79
8.1 RNA preparation.....	79
8.2 Compound Preparation.....	79
8.3 Unlabeled-AM1A control screen.....	80
8.4 5'-TAMRA-AM1A replicate screen.....	81
8.5 5'-TAMRA-RNA specificity screen and dose response assay.....	82
8.6 Ligand preparation for Computational Studies.....	83
8.7 Molecular Descriptors.....	83
8.8 Molecular Docking.....	84
8.9 Interaction Fingerprinting.....	84
8.10 Computational Development of AMC model and Grid.....	85
9. Calculations and Data Analysis.....	87
9.1 Collection of Region Interactions.....	87
9.2 Fluorescence Normalization.....	87
9.3 Error Propagation.....	88
10. Bibliography.....	89

## List of Table and Figures

<b>Table I</b> List of Abbreviations.....	14
<b>Figure 1.1</b> The T-box riboswitch.....	16
<b>Figure 1.2</b> Conserved Residues in the T-box antiterminator.....	17
<b>Figure 1.3</b> 5'-Dye bound antiterminator models.....	32
<b>Figure 2.1</b> Computational/Experimental project workflow.....	33
<b>Table 2.1</b> Reading plate setup – control assay.....	38
<b>Table 2.2</b> Reading plate setup – reproducibility assay .....	39
<b>Figure 2.2</b> Analysis example – reproducibility assay.....	40
<b>Table 2.3</b> Reading plate setup – specificity assay .....	41
<b>Figure 2.3</b> Analysis example – specificity assay.....	42
<b>Figure 3.1</b> FDA library computational process .....	43
<b>Figure 3.2</b> FDA library Lipinski violations .....	44
<b>Figure 3.3</b> Antiterminator regions bound by top FDA compounds.....	46
<b>Table 3.1</b> Candidate compounds from FDA library .....	47
<b>Figure 3.4</b> Poses and interactions of acebutolol, amodiaquine, and palbociclib .....	48
<b>Figure 3.5</b> Fluorescence control assay results – FDA compounds.....	49
<b>Figure 3.6</b> Reproducibility assay results – FDA compounds .....	50

<b>Figure 3.7</b> Specificity assay results – FDA compounds .....	51
<b>Figure 4.1</b> ZINC library computational process .....	53
<b>Figure 4.2</b> ZINC library Lipinski violations.....	54
<b>Figure 4.3</b> Charge distribution of druglike ZINC compounds .....	55
<b>Table 4.1</b> Candidate compounds from ZINC library .....	56
<b>Figure 4.4</b> Poses and interactions of 2,4-diaminobutyrate and l-phenylalaninamide.....	57
<b>Figure 4.5</b> Fluorescence control assay results – ZINC compounds .....	58
<b>Figure 4.6</b> Reproducibility assay results – ZINC compounds.....	59
<b>Figure 4.7</b> Specificity assay results – ZINC compounds.....	60
<b>Table 5.1</b> List of laboratory compounds .....	61
<b>Figure 5.1</b> Fluorescence control assay results – polyamines.....	63
<b>Figure 5.2</b> Reproducibility assay results – polyamines .....	63
<b>Figure 5.3</b> Specificity assay results – polyamines.....	64
<b>Figure 5.4</b> Fluorescence control assay results – polycyclic heterocycles.....	65
<b>Figure 5.5</b> Reproducibility assay results – polycyclic heterocycles .....	66
<b>Figure 5.6</b> Specificity assay results – polycyclic heterocycles.....	66
<b>Table 5.2</b> AM1A docking results – laboratory compounds.....	67
<b>Figure 6.1</b> Three-dimensional structure of AM1A and AM <sub>Control</sub> computational grids ....	68
<b>Figure 6.2</b> Docking comparison – Glide Emodel values.....	71

<b>Figure 6.3</b> Computational and Experimental results comparison.....	72
<b>Table 7.1</b> Compilation of thesis results by compound.....	73
<b>Figure 7.1</b> Initial dose response assays – amodiaquine and hydroxychloroquine .....	74

**Table I** List of abbreviations

Abbreviation	Definition	Abbreviation (Continued)	Definition
A	adenosine	MW	molecular weight
AARS	aminoacyl-tRNA synthetase	N	variable nucleobase
ASO	antisense oligonucleotide	NMR	nuclear magnetic resonance
C	cytosine	OPLS	Optimized Potentials for Liquid Simulations
DNA	deoxyribonucleic acid	OSC	Ohio Supercomputer Center
EDTA	ethylenediaminetetraacetic acid	PDB	Protein Database
EM	electron microscopy	QPlogS	predicted aqueous solubility
FDA	Food and Drug Administration	RFU	relative fluorescence unit
FRET	fluorescence (or Forster) resonance energy transfer	RNA	ribonucleic acid
G	guanine	RNase	ribonuclease
HTVS	High-Throughput Virtual Screen	SP	Standard Precision
K <sub>D</sub>	dissociation constant	TAA	tris(3-aminopropyl)amine
logP	partition coefficient	TAMRA	tetramethylrhodamine
MCE	MedChemExpress	tRNA	transfer RNA
MD	molecular dynamics	U	uracil
MMFF	Merck Molecular Force Field	ZINC	ZINC is not commercial (database)
MOPS	3-(N-morpholino)propanesulfonic acid		

## 1. Introduction

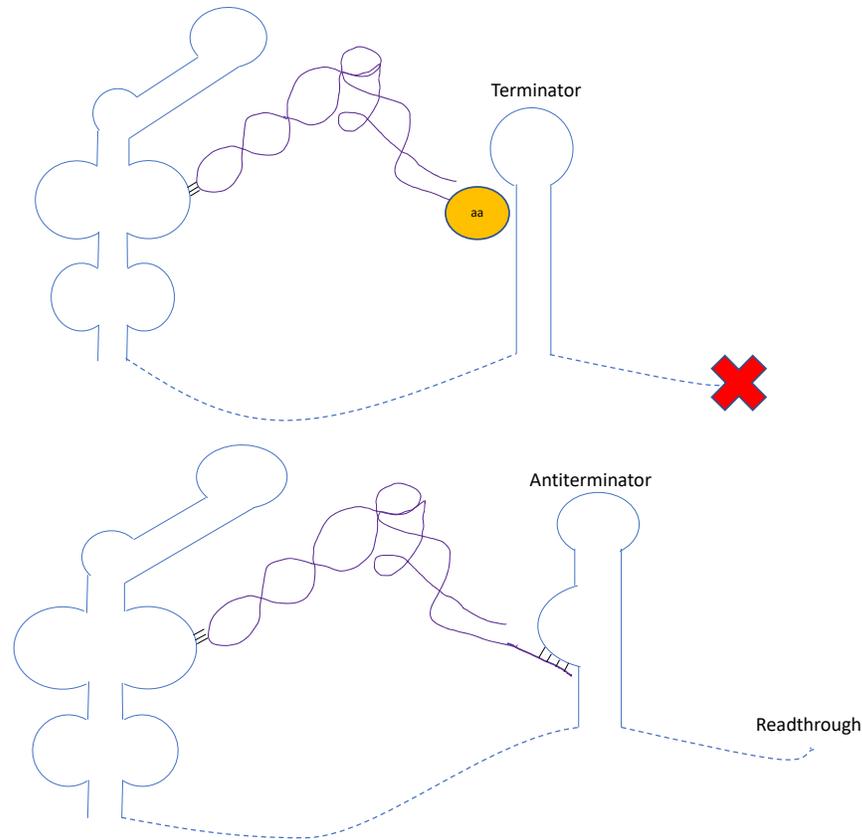
Antibiotic resistance is a growing risk to public health, afflicting two million in the United States annually (Centers for Disease Control and Prevention, 2018). As such, the search for novel treatments for these conditions has become more urgent. Within

bacterial RNA (ribonucleic acid, the functionalized, single-stranded equivalent of DNA), riboswitches are intriguing targets for drug discovery because they innately bind small molecules and are often linked with functions that affect the bacteria's survival (Deigan et al, 2011). Finding small molecules that disrupt the native function of riboswitches would be advantageous in antibiotic drug design as riboswitches are often well conserved and regulate multiple genes. One riboswitch structure that is highly conserved (its sequence is consistent) among infectious bacteria is the antiterminator sequence of the T-box riboswitch (Green et al., 2010).

### *1.1 The T-box Riboswitch*

Regulation of genes in response to environmental and metabolic stimuli is an essential function of noncoding RNAs such as riboswitches. The T-box riboswitch is an important 5'-end mRNA structure that determines expression of protein synthesis genes such as aminoacyl-tRNA synthetases (AARSs) – enzymes that combine amino acids to their cognate uncharged (unattached to amino acid) tRNAs - through transcriptional or translational control. The amino acid-bound (charged) tRNAs are the building blocks of protein synthesis. The T-box structure adopts two different conformations in its 3'-end dependent upon its binding to charged or uncharged tRNA, forming either the terminator or antiterminator structure (Figure 1.1). When attached to charged tRNA, the stabilized terminator helix prohibits downstream polymerization by RNA polymerase (transcription) or ribosomes (translation) that constructs AARSs. When attached to

uncharged tRNA, the kinetically formed antiterminator helix (Fig. 1) is stabilized and permits readthrough, thus synthesizing AARSs (Green et al., 2010).

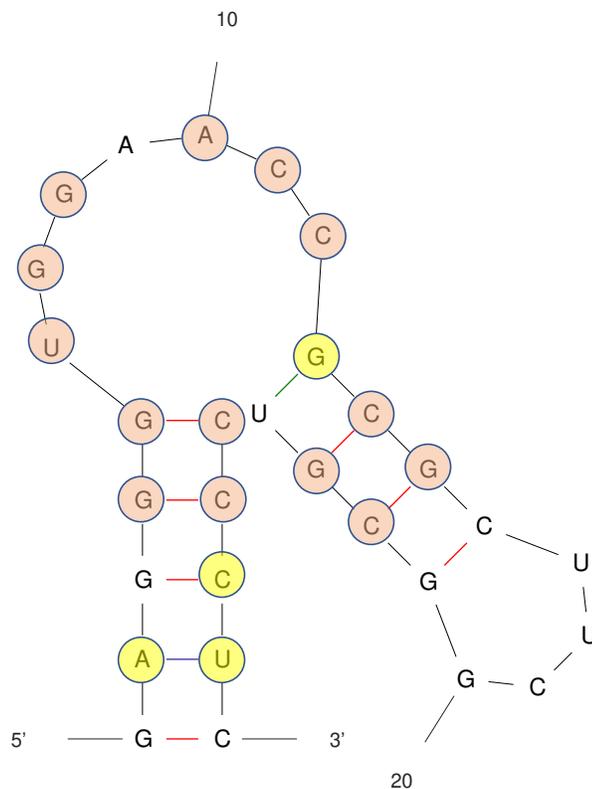


**Figure 1.1** The T-box riboswitch (blue) is responsive to the cellular ratio of charged and uncharged tRNA (purple), with antitermination resulting in readthrough stabilized by full tRNA-binding.

The antiterminator mechanism of the T-box riboswitch is highly conserved among Gram-positive bacteria (Deigan et al, 2011), meaning its sequence is maintained among a number of species such as the families *Streptococcus* and *Staphylococcus*, many of which can cause serious illnesses. Riboswitches can also be bound by non-native molecules such as antibiotic aminoglycosides (Aghdam et al., 2014), showing that selectivity of small molecules to the antiterminator is feasible (Fauzi et al., 2005). If the T-box can be bound at the antiterminator to block its native function, protein production would stall and survivability of infectious bacteria containing the T-box would be limited. The

prevalence and function of T-box riboswitches in infectious bacteria as well as their recurrence in multiple genes of the bacteria make them intriguing targets for novel drug development (Liu et al., 2015). In addition, since the T-box sequence is found mostly in prokaryotic organisms, the risk that drugs targeting this sequence would react with human nucleic acid sequences is greatly reduced.

### 1.2 Structure and mechanism of the T-box Riboswitch



**Figure 1.2** Primary sequence and secondary structure of the T-box antiterminator highlighting conserved sequences (Suddala et al., 2019) Highly conserved (>80% sequence identity) residues are illustrated in orange, and moderately conserved residues (50-80% sequence identity) residues are illustrated in yellow. Structure constructed in mFold web server.

The initial discovery and detailing of the T-box sequence characterized the *tyrS* gene T-box riboswitch of a hay bacteria, *B. subtilis* (Henkin et al., 1992). This T-box sequence controls the production of AARS enzymes which attached tyrosine to tRNA, allowing tyrosine to be attached onto growing proteins in the bacteria. Complementary T-

box riboswitches have been found for every amino acid in multiple bacterial species and have been identified using bioinformatic techniques such as comparative genomic analyses (Vitreschak et al., 2008). While parts of the overall sequence vary from species to species, many key regions of the riboswitch have been identified as being highly conserved, meaning part of the RNA sequence is retained in a high percentage of the species that have been analyzed (Figure 1.2) (Suddala et al., 2019).

Variations in the primary sequence (the sequential order of bases, e.g. UUAUCG) of RNA lead to alterations in the secondary structure (the folding of RNA upon itself from base pairing). However, even though the primary sequence of the T-box riboswitch is not completely conserved, there are two common modules of secondary structural motifs that exist in every T-box. T-boxes are comprised of two main regions bound together by variable linkers, those regions being the upstream Stem I and the downstream antiterminator/terminator domain (Zhang et al., 2015). The conservation of these two regions is essential for T-box function, as both regions are utilized in tRNA recognition and binding (Gutiérrez-Preciado et al., 2009).

Three-dimensional structure of RNA sequences provides information on the flexibility and conformation of the RNA and is found using techniques such as X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, and these structures can be modeled using bioinformatic approximation. Using NMR spectroscopy, both Stem I (Wang et al., 2010, and Wang et al., 2011) and the antiterminator (Gerdeman et al., 2003) have been characterized, and using crystallography Stem I bound to its corresponding tRNA has also been defined (Zhang, et al., 2013). More recently, some efforts have been made to model an entire T-box riboswitch-tRNA complex using X-ray

scattering analyses, and the resulting model shows the entire complex as a planar, circular region (Chetnani et al, 2017, and Fang et al., 2017), and further cryo-EM and cocrystal structures have revealed the riboswitch is composed of three regions that clamp tRNA (Battaglia, et al., 2019) and a central spine of tRNA to T-box stacking in riboswitches bound to uncharged tRNA lends the complex its stability (Li et al., 2019).

There are some questions and points of contention still remaining in structural studies of the T-box riboswitch. While there are models of the entire T-box-tRNA complex, these models are either considered ‘molecular envelopes,’ and are not resolved as definitively as NMR and crystallography experiments or are subject to discrepancies between the different structures. The total structure of the T-box leader region, unbound and in complex with tRNA, has yet be resolved.

Out of an array of molecules present within a cell, the T-box riboswitch effectively binds a corresponding tRNA to monitor the ratio of amino acid-bound and unbound tRNAs, turning the riboswitch off or on to transcribe AARS genes for protein production, respectively (Green et al., 2010). The mechanism for the specific binding of tRNA has been elucidated, as well as its role in antitermination/termination. The upstream Stem I region of the T-box contains a sequence called the specifier, which contains a single stranded segment that contains an anticodon, three bases complementary to the tRNA codon, a segment three single stranded bases that are unique to each tRNA type (Green et al., 2010). The specifier sequence is responsible tRNA identification by specifically binding the complementary tRNA to its correct T-box riboswitch.

The end of tRNA that binds amino acids has a conserved “acceptor” NCCA sequence, where N is a variable nucleobase (A,C,G,U) determined by the amino acid to

which the tRNA corresponds (Grundy et al., 1994). The acceptor end binds the T-box riboswitch downstream of Stem I, at the seven-nucleotide bulge of the antiterminator, which contains UGGN sequence at its 5' end that base pairs to the tRNA with N complementary to the relevant tRNA (Grundy et al., 1994). The discriminator base in tRNA also confers specificity to the antiterminator bulge, which supports the function of the bulge as another specifier in the T-box riboswitch (Grundy et al., 1994). Complexing of tRNA to the T-box system also affects the selection of antiterminator or terminator conformation (Gerdeman et al, 2002). Absent of intermolecular interactions, the terminator structure is thermodynamically favored over the antiterminator structure (Gerdeman et al., 2002, and Jentzsh et al., 2011). Binding tRNA overcomes this thermodynamic disadvantage to favor the antiterminator structure (Gerdeman et al, 2002). However, thermodynamic analysis indicates that base-pairing alone does not overcome the stability difference between the terminator and antiterminator (Hines, et al., 2010). Structural modification - hypothesized to be an induced fit, where binding of the tRNA alters the antiterminator structure- occurs in the antiterminator upon tRNA binding, and coaxial stacking of tRNA and the new tRNA-antiterminator helix assist in stabilizing the antiterminator to be favored over the terminator, allowing the production of AARS genes (Hines, et al., 2010).

As displayed by kinetic analyses, the binding of tRNA to the T-box complex occurs in two steps, the recruitment of tRNA by Stem I followed by stabilization of the antiterminator (Suddala, et al., 2018, and Zhang, et al., 2018). This two-step binding mechanism allows for the competition of amino acid-bound and -unbound tRNAs to both bind the T-box riboswitch, allowing the riboswitch to “monitor” the conditions of the cell

and only produce new AARSs when the ratio of unbound to bound tRNA is high, leading to more probable tRNA stabilization of the antiterminator structure.

### *1.3 Ligand identification and binding to the T-box riboswitch.*

While the native function of the T-box riboswitch is to bind corresponding tRNAs, a number of other molecules have been found to effectively bind regions of the T-box as well. Focusing on the antiterminator mechanism, the molecules that bind this RNA other than tRNA can be sorted into two classes: cofactors and drug candidate molecules. Cofactors are biologically relevant molecules that bind a riboswitch and are either requisite to its activity or enhances the activity rate. Drug candidate molecules are either naturally occurring or synthesized chemicals that may bind riboswitch and modulate its activity.

In early studies of molecular interactions with the T-box riboswitch, it was noted that a possible divalent (+2 charge) metal binding pocket could be present in the antiterminator mechanism of the T-box riboswitch (Means et al., 2005). Direct investigation into this possibility found that a diffuse magnesium binding site is present in a highly conserved region of the antiterminator (Jack et al., 2008), and that magnesium facilitates and improves the complete binding of the acceptor end of tRNA to the antiterminator, supporting an induced-fit model of tRNA binding as described above (Means et al., 2009).

Based on attempts to replicate tRNA binding to the T-box riboswitch *in vitro*, outside of a living cell, it is believed that an unknown protein cofactor is needed for proper binding to occur (Putzer et al., 2002). While the protein identity is still unknown, the polyamine spermidine has been shown to be an effective substitute by replicating

tRNA binding *in vitro* (Liu et al., 2018). At low concentrations, spermidine binds the antiterminator above the tRNA acceptor region and alters the flexibility of the region to encourage tRNA binding (Liu et al., 2018). Further research is needed to identify the protein that spermidine is emulating.

A number of synthetic and natural molecules have been assayed with the antiterminator mechanism to identify how they bind the antiterminator and what effects they have on antiterminator-tRNA binding. Aminoglycosides are a class of positively charged molecules that have previously been shown to bind ribosomal RNA. Furthermore, a study possible T-box antiterminator model RNA interaction with aminoglycosides that eight compounds in this class bound the antiterminator with micromolar dissociation constants (Means et al., 2005). It was also shown that the aminoglycosides bind the antiterminator through electrostatic interaction, with a possible mixture of specific and nonspecific binding, possibly binding in the bulge region. The difference in binding between the aminoglycosides also suggests a possible amenable RNA pocket or divalent metal binding site. However, it was later determined aminoglycosides do not have strong drug potential as they bind in the same area and manner as magnesium (Anupam et al., 2008). It was found to displace four monovalent ions suggesting the aminoglycoside binds in the divalent magnesium site on the antiterminator, as aminoglycosides often bind in metal-ion pockets in RNA. The location of the binding site was found to be at the 5' end of the bulge. Interestingly, it was also determined that the aminoglycoside neomycin-B was even found to enhance tRNA binding (Anupam et al., 2008). Studies on aminoglycosides altered the direction of drug discovery for the antiterminator towards identify molecules with little to no positive

charge, as solely electrostatic contacts could have nonspecific and/or stabilizing effects on the antiterminator.

Synthetic molecules designed to be favorable for binding to the antiterminator mechanism have been synthesized by the Bergmeier Lab at Ohio University (Means, et al., 2006, and Acquah-Harrison et al., 2010). These molecules have a triazole or oxazolidinone backbone and are designed to have side chains that could interact with the antiterminator through hydrogen bonds or hydrophobic interactions. Both oxazolidinone and triazole compounds have been identified in previous studies as capable backbones for RNA-binding drugs as (Hilimire et al., 2017, and Shaw et al., 2011). A series of combined oxazolidinone-triazole backbone compounds have been synthesized as well (Armstrong et al., 2020). Many of these compounds have been found to bind the antiterminator, and some have even interrupted tRNA binding and stabilization of the antiterminator (Zhou et al., 2012, and Armstrong et al., 2020). Research into these molecules is ongoing with new compounds being synthesized as well as binding behavior and inhibition activity being further characterized.

In addition to research on the antiterminator mechanism as a drug target, the Stem I region has also been a drug discovery target. This research has used the NMR-derived structure of the *B. subtilis* T-box Stem I as a computational model on which to simulate compound binding, and resulted in the discovery a number of molecules that tightly bound Stem I. Furthermore, these results were corroborated in vivo by blocking mRNA and translation of tRNA synthetase genes in bacterial assays, supporting the accuracy of computational screening (Frohlich et al., 2019).

#### *1.4 RNA drug discovery and design*

While developing therapeutics targeting RNA has been of interest for more than a decade, progress in identification and design of RNA inhibitors has lagged behind research ability to identify potential targets (Connelly, et al., 2016). Messenger and noncoding RNAs have highly structured elements, many of which are important to the function of the RNA (Warner et al., 2018). The unique space in such ordered regions is amenable to compound binding, which can be seen in the native function of certain non-coding RNAs such as riboswitches. The potential for RNA as a target for small molecule therapeutics is of growing interest, and the development of methods and the identification of important physicochemical properties which provide binding affinity and specificity will be of great importance for the development of this field.

A number of methods have been developed to identify small-molecule RNA binding. This includes computational and experimental approaches including biochemical screening, cell-based screening, structure-based docking, and sequence-based design (Shao, et al., 2020). Biochemical screenings screen compounds against an RNA target or targets with the ligand-target interaction indicated by a spectroscopic or chemical change. Spectroscopic approaches include fluorescence resonance energy transfer (FRET) (Means et al., 2005), fluorescent indicator displacement (Tran et al., 2012), and mass spectrometry (Rizvi et al., 2020). Such methods benefit from having scalability, capable of medium and high throughput assays and good accuracy but are sensitive to structural perturbation by their probes for fluorescence assays and can have limited sensitivity in the case of mass spectrometry (Yu et al., 2020). Cell-based screening techniques collect a phenotypic readout, such as reporter expression or cell survival as a response to

incubation with the compound of interest. This technique has been effective and efficient in the identification of seminal RNA-binding compounds such as Ribocil (Narasimhan et al., 2005), though it suffers from indirect identification of drug targeting which leaves more investigation to be performed to identify the molecular mechanism of the phenotypic response. Structure-based virtual screening methods make use of known three-dimensional structures of RNA, such as those determined in crystallographic, NMR, and cryo-EM studies, as a target for computational docking of libraries of small molecules. Computational dockings can be completed on a single conformer of the RNA or can be coupled with molecular dynamics (MD) to test the compounds against a dynamic ensemble of the RNA target (Shao et al., 2020). Molecular docking enables rapid sampling of a large variety of potential ligands but is hindered by low hit rates and the accuracy of computational predictions must be considered with scrutiny. Sequence-based approaches such as Inforna (Disney et al., 2016) consider the secondary-structural information of RNA targets rather than their three-dimensional structure. Known small molecule-RNA motifs are compiled in a library which is compared against a sequence of interest to provide a score of the fitness between the motif and RNA molecule. This can lead to rational drug design by selecting a strong pair or creating a modular ligand combining two binding partners with a linker, but the method is limited to the known binder database, so only certain sequences can be targeted effectively. In computational methods rational and fragment-based drug design can be employed to develop ligands as drug candidates (Shao, et al. 2020). Rational design is the approach to develop ligands using structural information of the target and using desired physicochemical properties to match the target, often developed and tested in a structure-activity series. Fragment based

design is a rapid method focused on testing many small groups and moieties against a target to identify high affinity groups which can be retroactively linked together into a larger selective molecule. Both methods rely upon the use of small compounds with desirable properties to target RNA.

Two main categories of chemicals have been employed as RNA-targeted drug design has developed: complementary oligonucleotides and druglike molecules. Oligonucleotide development of complementary or antisense oligonucleotides (ASOs) can bind and inhibit the translation of mRNA, as in the case of the retinitis oligonucleotide Vitravene, it can inhibit the activity of an RNA such as splicing (Spinraza, Nusinersen), or can engage in splicing activities by engaging endogenous siRNA pathways (Juru et al., 2020). Although ASO-drug development is a promising expansion of RNA-drug targeting, there are significant challenges to their adoption and widespread use. Owing to their size and large negative charges, ASOs struggle in delivery into cells (Shao et al., 2020) as well as non-hepatic tissues and are unable to cross the blood-brain barrier unless directly injected into the spinal canal (Juru et al. 2020). ASOs are also subject to catabolism by RNases (Yu et al. 2020) and can cause immunological responses (Juru et al., 2020).

Drug-like molecules are less inhibited by their chemical qualities, tending to be orally available and with good cellular mobility. Molecules are said to be “drug-like” if they have physicochemical properties that align well with previously identified therapeutic chemicals, providing a shorthand for their potential to share desirable properties with these chemicals (Warner et al., 2018). One of the most commonly used metrics for tagging molecules as drug-like, developed from retrospective analysis of

approved therapeutics and drug candidates, is Lipinski's Rule of Five (Lipinski, 2004). The parameters for the Rule of Five are that the compounds have a molecular weight lower than five hundred, have a partition coefficient (logP) less than or equal to five, have up to five hydrogen bond donors, and as many as 10 hydrogen bond acceptors. At the time of the inception of the Rule of Five, 90% of orally active drugs achieving phase II clinical status aligned with these benchmarks (Lipinski, 2004). As a consequence of the historical priority of protein-targeted drug design, the standard for drug-likeness is mostly characterized by a field of compounds optimized to associate with peptides, and it is possible that RNA-binding compounds have different structural and kinetic properties that would violate traditional drug-like standards (Warner et al., 2018). However, all compounds considered for drug design are still governed by the constraints of bioavailability, and as such drug-likeness remains a strong metric to consider in RNA drug discovery research.

A continued challenge in the drug design of RNA-binding molecules is specificity and selectivity (Connelly et al., 2016). A difficult task in producing biomolecule targeting compounds is to tune the binding specificity of the candidate compound to avoid binding undesirable targets (narrow selectivity) while allowing for coverage of the target of interest (broad selectivity) (Huggins et al., 2012). The aim of drug development is to identify small molecules that strongly binds its target but does not bind to other similar macromolecules as to interfere with body functions in the host (Ryde et al., 2016). This is pronounced in RNA-targeted drug discovery, where it is necessary to achieve high binding affinity and selectivity with the challenge of targeting a macromolecule composed of 4 possible monomeric units with many similar RNA motifs within a cell

(Warner et al., 2018). Aminoglycosides offer effective broad-spectrum antibiotic activity through their binding of ribosomal RNA, but this limits their ability as a scaffold as their selectivity between different RNAs is low (Juru et al., 2020), and ribosomal targeted RNA-binding compounds are considered a poor training set for drug development due to the uniquely high concentration of their target in cells (Warner et al., 2018). Additionally, highly basic and intercalating compounds also offer poor prospects for selectivity as they interact with phosphates and nucleobases respectively, both of which are essential components to every monomeric unit within an RNA molecule (Warner et al., 2018). As such, the available space for ligands targeting RNA with high specificity is constrained both their physical properties as well as the scaffolds upon which they are developed.

In a study of previously identified RNA-targeted bioactive ligands and FDA-approved small molecule drugs, it has been found that such RNA-binding compounds often exist in a unique physicochemical space: compliance to medicinal chemistry rules and unique structural groups and overall shape (Morgan et al., 2017). Compared to FDA-approved small molecules, which mostly target proteins, RNA-binding ligands have been found to have a greater content of nitrogen and rings, specifically heterocycles, while containing fewer oxygens,  $sp^3$  hybridized carbons, and stereocenters (Morgan et al. 2017). This original study was expanded and developed into a community accessible library, and chemoinformatic analysis of this larger database of bioactive RNA ligands again indicated unique physicochemical traits of RNA ligands compared to FDA compounds (Morgan et al., 2019). In these studies, it was found that both monovalent and multivalent RNA-binding compound tend to have a positive charge per every 250-350 atomic mass units, indicating a significant role charge and electrostatics play in ligand-

RNA interaction (Morgan et al., 2017). It has been noted that RNA-binding ligands have higher topological polar surface areas and positively charged surface areas compared to other drug sets (Aboul-ela, 2010). Other RNA-targeting studies have supported the physicochemical properties present in RNA-binding small molecules such as an expanded number of heteroaromatic rings (Rizvi et al., 2019) as well as a greater fraction of aromatic atoms and a lower fraction of  $sp^3$  carbons compared to FDA compounds (Haniff et al., 2020). Small molecules are that they appear more likely to assume a rod-like shape than FDA compounds, and in turn are less likely to be sphere-like (Morgan et al., 2017). The possible privileged RNA-binding space indicated by such physicochemical traits in known RNA binders has fueled the scaffold-based drug design approach, which focuses on optimization of previously recognized moieties to target a specific RNA structure (Juru et al., 2020). Identification of these recognized structures can be completed through many pathways, such as adaptation of DNA-binding molecules (Zapp et al., 1997), or high throughput virtual screening of RNA dynamic ensembles (Stelzer et al., 2011). Numerous scaffolds have been identified and investigated including aminoglycosides, oxazolidinones, imidazole, and aminoquinolones (Juru et al., 2020), and opportunity remains to expand this collection with further study of RNA-molecule binding.

Previous research in T-box antiterminator drug discovery has used rational drug design to develop structure-activity series using known RNA-binding scaffolds, and specificity has been addressed by testing ligand effects (binding, inhibition) against models of the antiterminator with similar sequences (Means et al., 2006). A bioinformatic study was completed to investigate thermodynamic differences between the

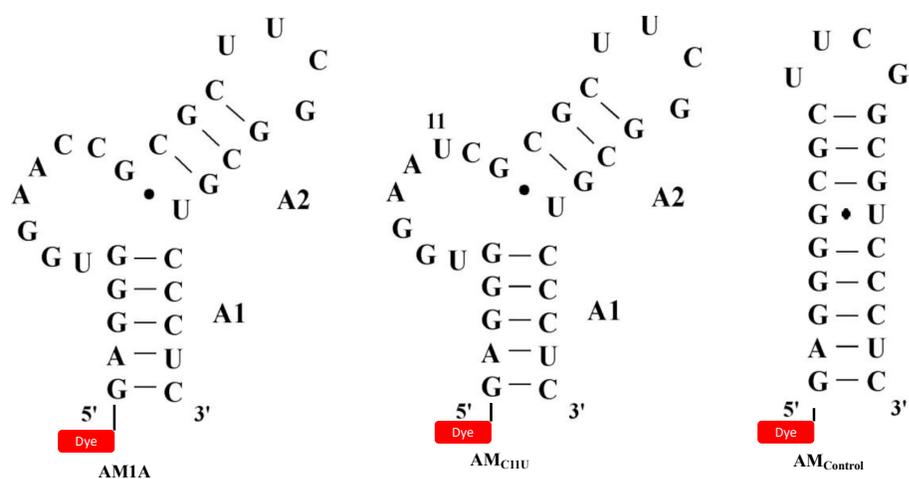
T-box riboswitch terminator and antiterminator structures across a range of different target AARS genes in different bacterial species (Jentsz et al., 2011). It was found that the free energy difference was similar across all species, but that glycyl T-box riboswitches might be better targets for drug development as the energy difference between terminator and antiterminator is great enough that a molecule could bind the antiterminator to preclude tRNA binding but not prevent the formation of terminator through stabilization (Jentsz et al., 2011). The antiterminator structure forms prior to tRNA interaction with the T-box and structural rearrangement may occur in the antiterminator upon tRNA binding, which would support an induced-fit model of this interaction (Green et al., 2010). These two modes of antiterminator structure and the thermodynamic gap between the terminator and antiterminator support the development of drugs to target the antiterminator and preclude tRNA binding or destabilizing the antiterminator-tRNA complex.

Previous research into the antiterminator selection of modified tRNAs showed that the binding of functionally active sequence AM1A fell into two categories, functionally relevant acceptor binding and a kissing loop interaction, while the reduced function variant, AM<sub>C11U</sub>, showed only the functionally relevant interaction (Fauzi et al. 2005). The results indicated the in vitro selectivity for the acceptor end of tRNA by the T-box antiterminator and that there are structural and dynamic components to selection, as shown by the different selectivity identified between AM1A and AM<sub>C11U</sub>. This suggests that specificity can be determined between antiterminator RNA models. The potential to drug the antiterminator led to the synthesis and experimentation to test the binding characteristics of a small library of 4,5-disubstituted oxazolidinones (Means et

al., 2006). To investigate the ability of these compounds to overcome the challenges binding affinity and specificity pose to RNA drug discovery, FRET experiments were completed using AM1A and AM<sub>C11U</sub>, two similar two model antiterminator RNAs differing only at a single nucleotide polymorphism. The investigation of ligand-RNA binding and specificity could then be evaluated by comparing fluorescence modulation of the two models in the presence of the oxazolidinones. Two compounds showing large differences between modulation of AM1A and AM<sub>C11U</sub> structures, validating the possibility of creating compounds that are highly specific to the T-box antiterminator (Means et al., 2006). An additional fluorescence method that has been employed to indicate specificity differences in antiterminator-ligand interaction is the use of 5' end single-dye bound RNAs in a steady state fluorescence assay (Means et al., 2007). This method, also known as a terminal target labeling assay, operates with the supposition that ligand binding influences the conformation and dynamics of the bound RNA, such that the bound dye experiences dose-dependent fluorescence modulation (Hermann, 2016). This method was expanded to include a third model, known as AM<sub>control</sub>, which does not include the acceptor-binding seven-nucleotide bulge of the antiterminator to indicate if a ligand may have an effect on the bulge region, yielding moderate binding site information in a relatively straightforward assay.

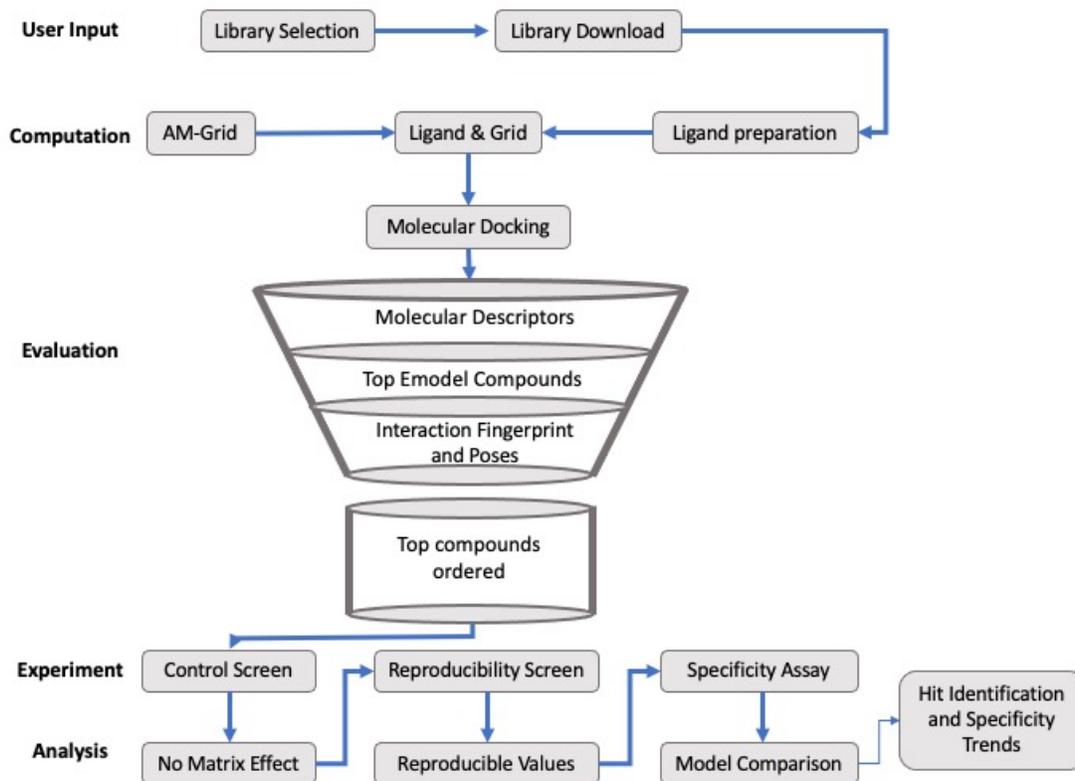
Computational studies have also been employed to develop theories on the compound binding sites to the antiterminator structure using the NMR-derived solution structure (PDB:1N53) using molecular docking and dynamics simulations (Liu et al., 2016). The use of computational and fluorescence methods has been developed to provide a primary screening of potential ligands for the T-box antiterminator. In an effort

to preliminarily characterize the specificity of ligands for the antiterminator and to provide key structural insights for development of a pharmacophore model. These two disparate methods, computational molecular docking and a terminal target labelling assay, were combined into a hybrid method that investigates theoretical energetic differences and fluorescent modulation of the binding of ligands to three models of the T-box antiterminator, AM1A, AM<sub>C11U</sub>, and AM<sub>Control</sub> (Figure 1.3).



**Figure 1.3** Secondary structure and sequences of 5' terminal dye-bound antiterminator models AM1A, AM<sub>C11U</sub>, AM<sub>Control</sub>.

## 2. Screening Design



**Figure 2.1** A project workflow was devised for the combination assay

In preparation for this hybrid study, a project workflow was devised to integrate computational molecular docking into the moderate-throughput fluorescence terminal target labelling screening (Figure 2.1). Each section in bold is expanded upon below as a discussion of the important aspects and principles of each step in the combination assay.

### 2.1 Computational Setup and Evaluation

**2.1.1 User input:** Compounds from research groups, papers, or libraries are selected according to selected parameters such as having druglike features, containing select functional groups, being a select size or weight, or being similar to other molecules. If these libraries are available in a Maestro (Schrödinger) readable file format,

they will be downloaded for the computational procedure. If the molecules need to be constructed, they will be created and have their structural energy minimized using the basic minimization tool (MMFF) in Spartan 10 and exported in .mol2 file format for further use. An example of a previous library identified are compounds from literature produced by Dr. Mir Hosseini (University of Strasbourg) that are aliphatic and contain aromatic amines showing structural similarity to either polyamines or histidine-like compounds. In this study, the MedChemExpress (MCE) FDA-Approved Drug Library, a collection of 2483 marketed drugs, and the ZINC natural compounds (natural-products + for-sale + named), containing 2962 compounds, were selected (Irwin et al., 2003). Both libraries provided their own .sdf file, which is compatible with Schrödinger software.

*2.1.2 Computation:* The compounds are processed through the OPLS\_2005 force field – a coverage enhanced version of OPLS\_2001 (Kaminski et al., 2001) – using the ligprep program in Maestro (Schrödinger) to develop all possible states of the compounds at the pH conditions of experimental screening –  $\text{pH}=6.5\pm 0.2$  for the terminal target labelling assays - and produce all possible isomers/tautomers of the compounds. The compounds may additionally be analyzed using Molecular Descriptors, a program which identifies properties of the ligands with its own terms as well as QikProp terms, such as molar mass or violations of Lipinski's Rule of Five. An antiterminator receptor grid has been developed using the NMR-solution structure of *B. subtilis* TyrS (PDB:1N53; Gerdeman et al., 2003), which has been optimized using the Glide module in Schrödinger. This model does not contain the UUCG tetraloop at the top of the A2 helix, since it is not a part of the conserved element of the antiterminator. In addition, the antiterminator is mutated at variable bulge nucleotide from U to A to match the

experimental functional model AM1A. There is also an ongoing effort to develop homology models of AM<sub>C11U</sub> and AM<sub>Control</sub>- two variants of the antiterminator with slightly different sequences, a cytosine to uracil substitution at the 11<sup>th</sup> position and deletion of bulge region, respectively - using Molecular Dynamics simulation. The compounds and grid (AM1A, and possibly AM<sub>C11U</sub> and AM<sub>Control</sub> as well) are added to the same project as the ligands and then computational molecular docking is completed, which will show optimal locations for molecule-RNA interactions and also give an output of estimated interaction energies and distance from residues, which can be exported as a .csv file. The docking is completed using the Glide program (Freisner et al, 2004) in Maestro with either the standard precision protocol (SP) or using high-throughput virtual screening (HTVS). Glide (grid-based ligand docking with energetics) docking uses exhaustive search techniques to “funnel” ligand conformations through a number of tests and refinements and can be further optimized following conformational refinement with a minimization using Monte Carlo (random sampling) algorithms. It then reports scores for ligand docking poses, including the Schrödinger software proprietary values GlideScore and Glide Emodel. Glide Emodel is a useful predictive binding score, as it combines the energy-grid score, GlideScore, and internal strain energy, yielding the most accurate prediction of binding affinity for the program. The number of optimal poses output for each ligand can be selected. Having multiple poses can be used to identify the consistency of binding and energy trends for each ligand, provide a qualitative metric of specificity in the docking procedure. Aforementioned computational chemistry procedures in the combination assay were completed at the Ohio Supercomputer Center (OSC).

Following docking, the output file (.maegz) was imported to a personal computer and was processed through the interaction fingerprints program in the Discovery and QSAR module in the academic licensed Maestro. This provides a count of ligand interactions to residues, similar to per residue interaction scores from the docking, but provides a quick score of the number of favorable interactions a ligand has to each ligand based off of empirical values.

*2.1.3 Evaluation:* After docking simulations have been completed, compounds are selected for fluorescence experimentation if they meet chosen criteria of strong binding scores, druglike properties, low charge and consistently binding in one region of the antiterminator according to interaction fingerprinting (A1, A2, bulge). Additional compounds were selected that spanned across the regions if they had exceptionally strong binding Emodel values. Charge was an important initial criterion for selection. Molecules with little to no positive charge are better candidates for specific binding to the antiterminator, as compounds that bind electrostatically (through charged elements) to RNA can be nonspecific (Liu et al., 2016) and electrostatic interactions have been found to either not affect or enhance tRNA binding to the T-box antiterminator (Anupam et al., 2008).

Some other criteria that are likely to be used are Glide Emodel, pose consistency, molar mass, and interaction number. The Glide Emodel is a composite score from computational docking that indicates the favorability of a compound-target interaction pose and is used to indicate strength of interactions (Freisner et al, 2004). Compounds are selected based on the more stringent of two cutoff criteria, those being either the top 10% of Emodel scores or scores less than or equal to -100 kJ/mol. The poses of each model

interaction indicate specificity of binding, as it can be determined if the compound consistently binds to a similar area of the antiterminator and with similar interaction types. This can be evaluated quantitatively by comparing the fingerprint interaction scores of each antiterminator region. Compounds that have only two distinct binding positions or consistent placement of all poses were selected for further evaluation. As indicated in Lipinski's Rule of Five, drug-like molecules often contain certain molecular properties, and as such compounds were filtered out if they showed in rule of five violations according to their molecular descriptors output. Interactions strength by type are estimated and reported in the docking spreadsheet as per residue (nucleobase) interaction scores by interaction type, such as salt bridges, hydrogen bonds, and pi stacking. In conjunction with pose information, the scores will help to determine predicted bond locations. Additional criteria that can be considered are selecting molecules based off of their shape (Morgan et al., 2017) or interaction type (Padroni et al., 2020). Compounds that fulfill all requirements of computational evaluation were selected and purchased if commercially available. Upon arrival, the compounds were prepared for fluorescence assays via dilution to a micromolar concentration and were stored in conditions in accordance with their MSDS in preparation for the fluorescence experiment series.

## *2.2 Experimental Design and Analysis*

*2.2.1 Fluorescence Experimentation (Control screen):* A control screen for autofluorescence or matrix effects on the unlabeled antiterminator RNA was completed for each ligand tested. This assay was completed for at most 2 ligands per procedure, with the following conditions in each reading well of 384-well microplate: 55 mM MOPS buffer, pH= 6.5, 50 mM NaCl, 0.01 mM EDTA. In addition, a row was tested which contained 100 nM unlabeled AM1A to evaluate if there were any fluorescence effects in the presence of the RNA. Test wells contained 10  $\mu$ M of the ligand. Control and treated wells were first prepared in a mixing plate and then 20  $\mu$ L of each mixture were transferred to a microplate (Table 2.1) and read in a Spectramax 5 fluorimeter.

**Table 2.1** Reading plate loading procedure for control screen. X denotes empty well, L(#) denotes a selected compound.

	1	2	3	4	5	6	7	8	9	10	11	...	18	19
A	X	X	X	X	X	X	X	X	X	X	X	...	X	X
B	X	MOPS Control 1	x	MOPS Control 2	x	MOPS Control 3	x	L1	x	L1	X	...	L2	X
C	X	X	X	X	X	X	X	X	X	X	X	...	X	X
D	X	Unlabeled AM1A Control 1	x	Unlabeled AM1A Control 2	x	Unlabeled AM1A Control 3	x	L1	x	L1	X	...	L2	X
E	X	X	X	X	X	X	X	X	X	X	X	...	X	X

The reading wells were placed in alternate rows and columns as it was found in previous unpublished research in the Hines Lab that the Spectramax 5 fluorimeter systematically fluctuates fluorescence readings between adjacent columns and rows.

*2.2.2 Analysis (Control Screen):* The control screen was used to identify whether the tested compounds would present other fluorescence effects in the following assays. As such, if the fluorescence of the control and test wells had similar RFU values to the blank wells (usually less than 0.5) then the compound would be tested in the reproducibility assay.

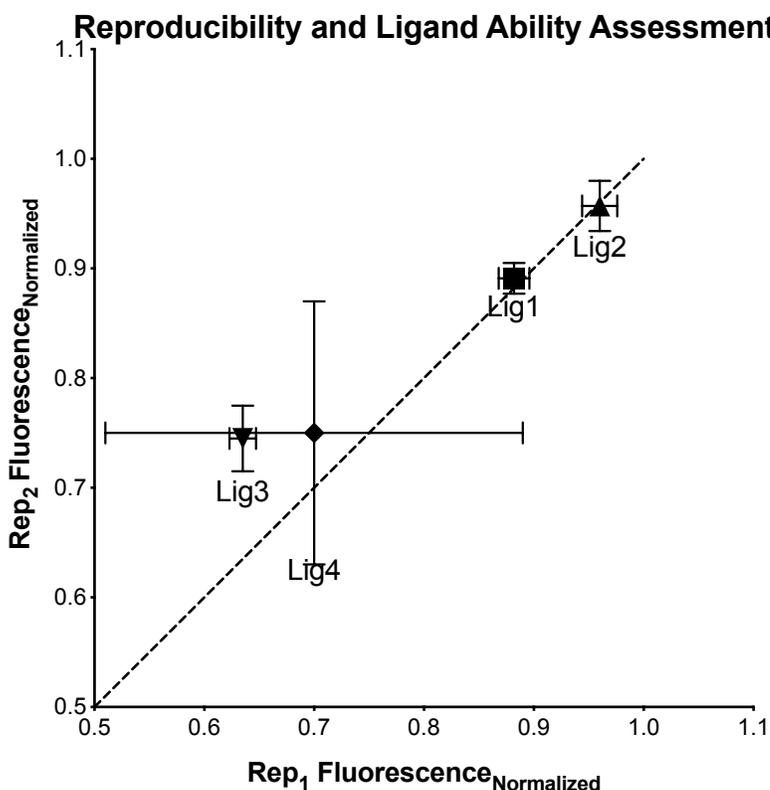
**2.2.3 Fluorescence Experimentation (Reproducibility screen):** A primary fluorescence screening of potential ligands was completed with AM1A in a steady-state fluorescence assay (Liu et al., 2015). The conditions for the assay were: 55 mM MOPS buffer, pH= 6.5, 50 mM NaCl, 0.01 mM EDTA, 100 nM 5'-tetramethylrhodamine (TAMRA) labeled AM1A. Experimental wells additionally contained 10  $\mu$ M of the compound being tested. Control and treated wells were placed in a microplate (Table 2.2) and read in a Spectramax 5 fluorimeter. A replicate row was then assayed in a second iteration of preparation.

**Table 2.2** Reading plate loading procedure for primary screen. X denotes empty well, L(#) denotes a selected compound. Rows B and D would have been completed at separate times.

	1	2	3	4	5	6	7	8	9	10	11	...	18	19
A	X	X	X	X	X	X	X	X	X	X	X	...	X	X
B	X	Control 1	x	Control 2	x	Control 3	x	L1	x	L1	X	...	L2	X
C	X	X	X	X	X	X	X	X	X	X	X	...	X	X
D	X	Control 1	x	Control 2	x	Control 3	x	L1	x	L1	X	...	L2	X
E	X	X	X	X	X	X	X	X	X	X	X	...	X	X

**2.2.4 Analysis (Reproducibility Screen):** For the primary screening the mean fluorescence for each ligand+AM1A well were normalized against the mean of controls in its corresponding row (Rep1/Control1). The replicate screenings were graphed with the propagated error as the first replicate on the x axis and the second replicate on the y axis (Figure 2.2). This method provides a quick visual indication of reproducible results, as those with high reproducibility fall on the line of identity ( $x=y$ ). In Figure 2.2, compounds 1 and 2 (L1 and L2) show high reproducibility and would be selected for the secondary specificity screening. However, only compounds with significant deviation

from zero have a strong effect on the antiterminator's structure, which means compound 2 may not be a strong binding candidate. Compounds 3 and 4 (L3 and L4) would not be selected due to unacceptable deviation from the line of identity or large error bars respectively.



**Figure 2.2** Reproducibility and ligand ability indication can be quickly assessed by comparing replicate value against a line of identity. Normalized fluorescence values were used.

### 2.2.5 Fluorescence Experimentation (Specificity Screen) The secondary

specificity assay of intriguing targets selected from the primary assay are tested under the same reaction conditions as described in the primary screening, but also with other 5-TAMRA labeled AM models:  $AM_{\text{control}}$  and  $AM_{\text{C11U}}$ . Each model is tested once, in a single row with up to two ligands.  $AM_{\text{control}}$  does not contain the functionally relevant bulge region of the antiterminator, and  $AM_{\text{C11U}}$  and have reduced functionality compared

to AM1A due to a C to U transition at the 11<sup>th</sup> position, which is in the bulge of the antiterminator.

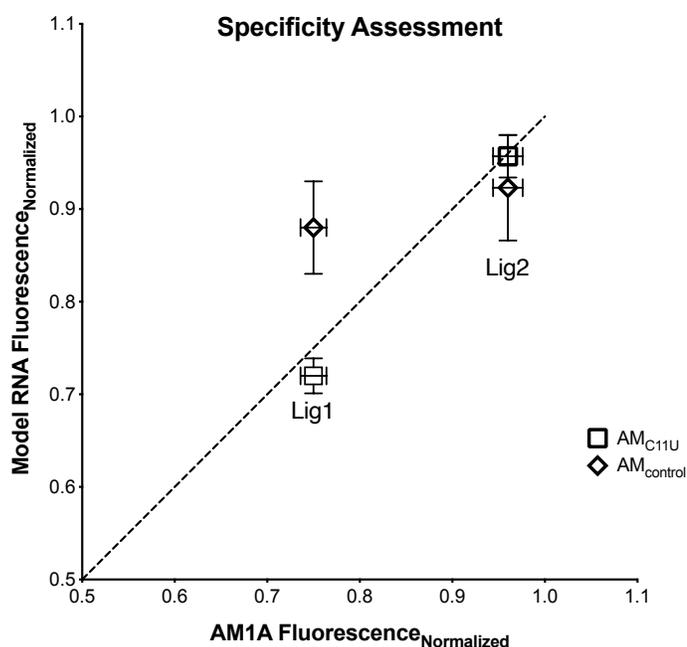
**Table 2.3** Reading plate loading procedure for secondary screen. X denotes empty well, L(#) denotes a selected compound. Each model RNA is tested prepared and tested separately.

	1	2	3	4	5	6	7	8	9	10	11	...	18	19
A	x	x	x	x	x	x	x	x	x	x	x	...	x	x
B	X	Ctrl AM1A	x	Ctrl AM1A	x	Ctrl AM1A	x	L1 AM1A	X	L1 AM1A	X	...	L2 AM1A	X
C	x	x	x	x	x	x	x	x	x	x	x	...	x	x
D	X	Ctrl AM <sub>C11U</sub>	X	Ctrl AM <sub>C11U</sub>	X	Ctrl AM <sub>C11U</sub>	X	L1 AM <sub>C11U</sub>	x	L1 AM <sub>C11U</sub>	X	...	L2 AM <sub>C11U</sub>	X
E	x	x	x	x	x	x	x	x	x	x	x	...	x	x
F	X	Ctrl AM <sub>Control</sub>	X	Ctrl AM <sub>Control</sub>	X	Ctrl AM <sub>Control</sub>	X	L1 AM <sub>Control</sub>	X	L1 AM <sub>Control</sub>	X	...	L2 AM <sub>Control</sub>	X
	x	x	x	x	x	x	x	x	x	x	x	...	x	x

*2.2.6 Analysis (Specificity Screen)* For the secondary screening, the mean normalized fluorescence and propagated error of each compound were compared between models. AM<sub>Control</sub> and AM<sub>C11U</sub> were compared against AM1A, the functionally relevant model (Figure 2.3). AM<sub>Control</sub> does not contain the bulge region crucial to the function of the antiterminator and therefore its comparison to AM1A indicates the importance of the bulge in structural modulation by compound-RNA binding. AM<sub>C11U</sub> contains a similar bulge to AM1A, so this provides a narrower view of ligand specificity. Compounds with significantly different normalized fluorescence values between models (not lying within error on the line of identity) were considered as showing binding specificity between model RNAs, and further research may be conducted in Hines Lab to elucidate their binding affinity (associated with  $K_d$ ) and effect on the antiterminator mechanism.

The results of the analysis of experimental data were then compared against the docking experiment results to identify which compounds showed agreement in binding

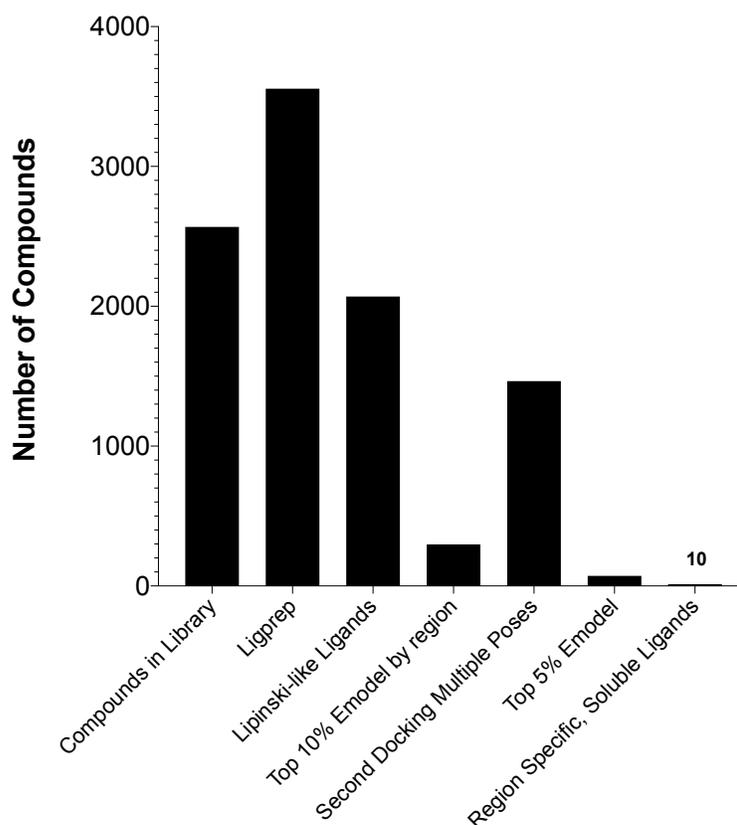
energies to fluorescence modulation and pose location to modulation of the different antiterminator models. Trends in functional groups, important residue interactions or types of molecules whose computational and experimental results both show favorable binding and specificity were then identified. These areas of agreement can serve as parameters for searching other libraries of compounds or to synthesize molecules that contain effective elements identified in the assay (combinatorial synthesis). Effectively, the results of one round of experimentation and analysis can help inform what parameters should be set for the next round, yielding an iterative process that will more effectively identify antiterminator-specific binding compounds with each cycle.



**Figure 2.3** Specificity is assessed comparing normalized fluorescence values between antiterminator models. Error bars are propagated error values.

### 3. Results – FDA-Approved Library

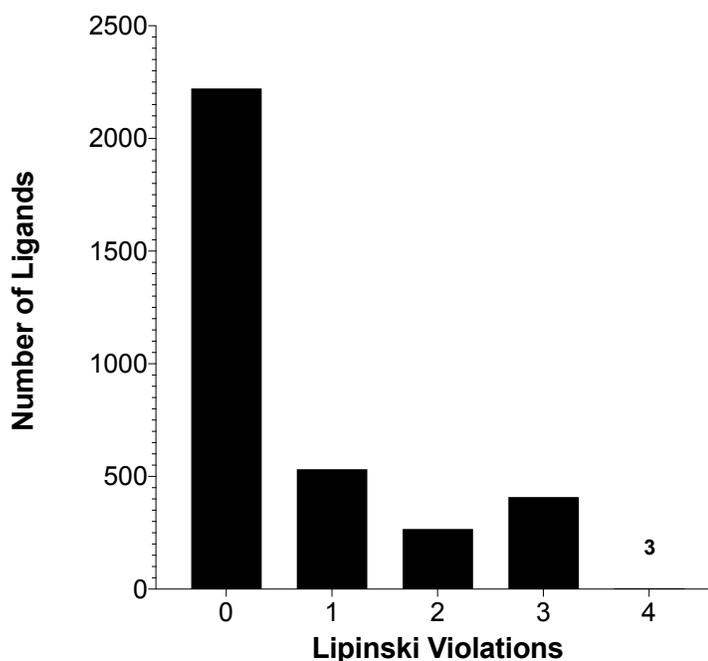
3.1 Computational preparation and docking yields potential antiterminator ligands in an FDA drug-repositioning library with high interaction scores and positional consistency.



**Figure 3.1** The computational process to identify strong ligand candidates to the T-box antiterminator identified ten ligands out of the original MCE FDA library.

The MedChemExpress, abbreviated to MCE, (New Jersey) FDA-Approved Drug Library Plus is a commercially available collection of compounds previously approved for therapeutic use as both a physical compound library and a computational library as a spatial data file (.sdf). The compounds in the library have a wide range of targets, including bacteria, protein channels, fungi, viruses, and protein receptors. The MCE library was used in the initial implementation of the hybrid assay workflow (Figure 3.1). The data file for the library compounds was collected and three-dimensional models of the compound, including tautomers and stereoisomers, were generated using LigPrep

program in Maestro with the OPLS\_2005 force field in pH conditions  $7.0 \pm 0.5$ . The LigPrep program yielded 3556 total potential ligands. However, 40 compounds amounting to approximately 1.6% of the original library, were dropped during the LigPrep process, not returning any structure which is common in computational chemistry library studies (Brooks et al., 2011).



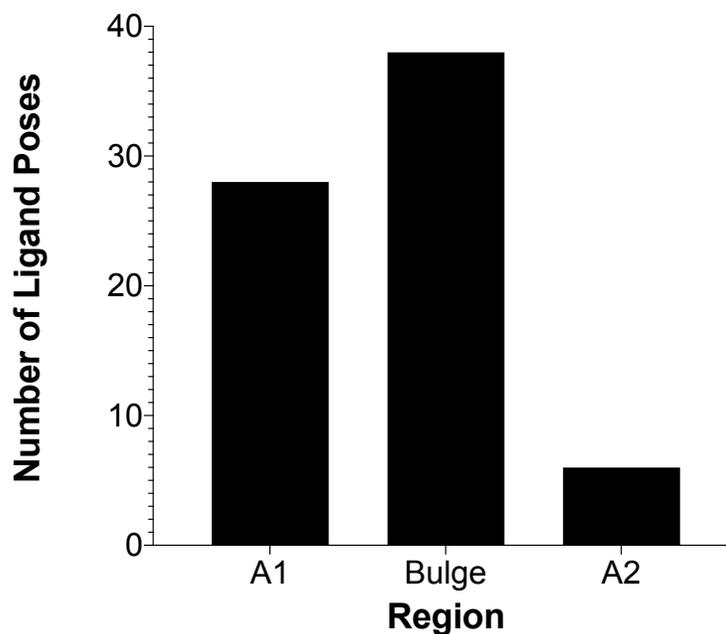
**Figure 3.2** MCE compounds mostly fell within the bounds of Lipinski's rule of five, but some compounds did not, an indication of the potential of therapeutics to exist outside of druglike space.

Following the construction of the three-dimensional library of compounds in experimental conditions, the molecular descriptors for the compounds were collected using the Molecular Descriptors program in Maestro. Compounds violating Lipinski's rule of five were removed, reducing the library down to 2070 compounds, representing a vast majority of the original library. As this is a library of previously approved FDA compounds, it is unsurprising that most compounds fit within the parameters of the rule of five (Figure 3.2). Interestingly, 3 compounds, fidamoxicin, talaporfin, and tenapanor, were found to have violated all parameters of the rule of five, as they are extended

molecules and macrocycles with multiple hydrogen bond donors and acceptors. This is still a large library to complete a multi-pose docking simulation to the antiterminator, so the compound library was first docked to the grid of the antiterminator with only a single pose per ligand using the high-throughput virtual screening (HTVS) protocol in Glide docking program in Maestro. The compounds with the lowest 10% of Emodel values, split to collect the top 3.3% per region of the antiterminator (A1, Bulge, and A2) as determined by the Interaction Fingerprint program in Maestro, were selected for further multi-pose ligand docking. Some compounds had multiple enantiomers or protonation states present in the pH conditions of the simulation, so these states were included as well to give a holistic representation of potential poses, yielding 297 compounds overall to be tested in the multi-pose docking.

A Glide docking of these top compounds was completed, recording up to 5 poses per ligand, which yielded 1465 total poses. Interaction counts were again generated for these poses using the Interaction Fingerprint program, and the total interactions for each antiterminator region were again used to sort poses into by the highest interaction totals for each region. Compounds with the most negative Emodel value, indicating the best interaction energy, of binding to the antiterminator for each compound, were then evaluated for consistent poses. This was completed by collecting the top 5% of poses by Emodel, annotated with the region for which they appear to be selective, and determining if more than one additional pose appears with to have more interactions in one of the other regions. The majority of these poses had the most interactions with the bulge,

numbering 38, while 28 interacted mostly with A1, and 6 interacted mostly with A2 (Figure 3.3).



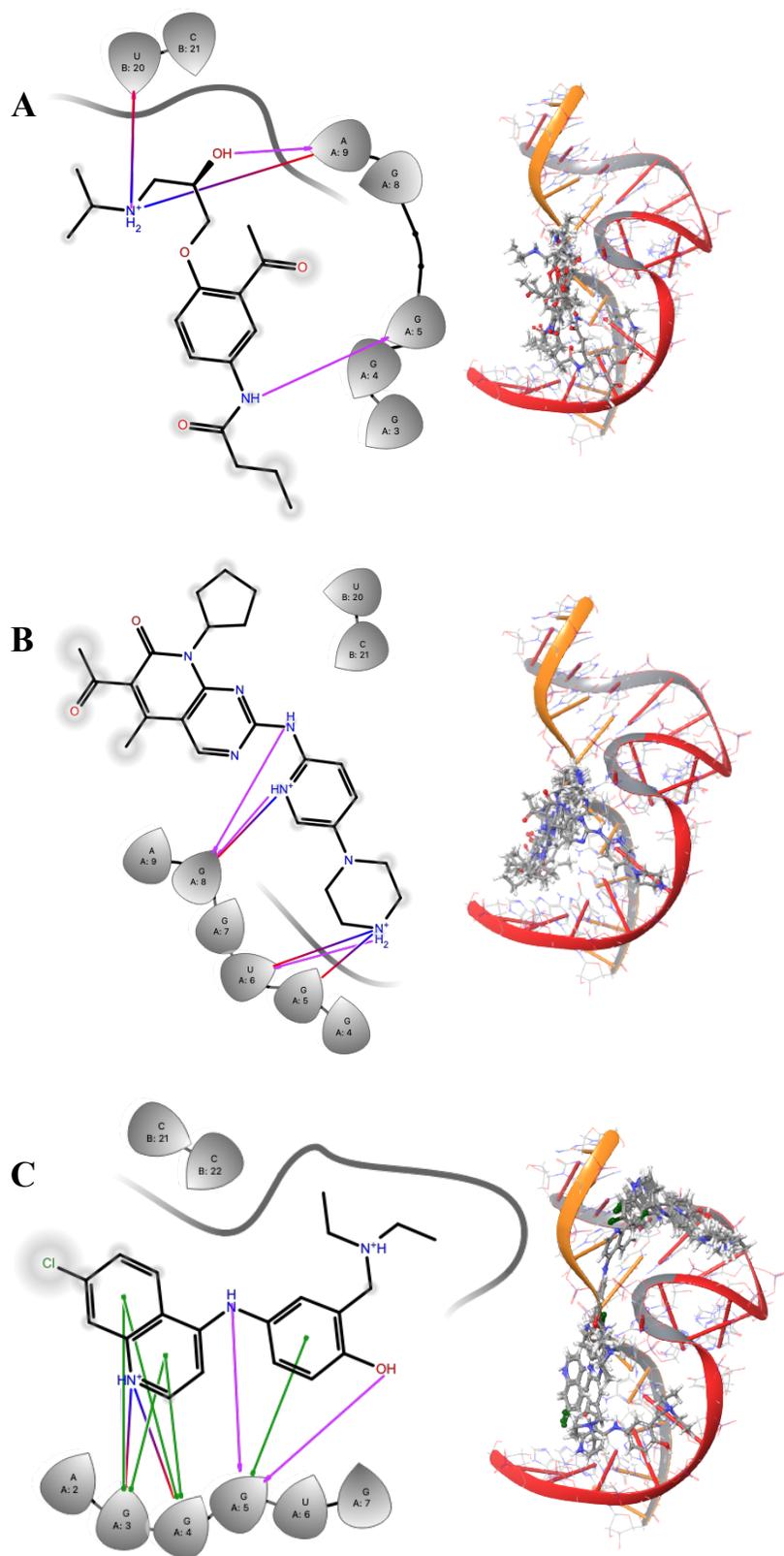
**Figure 3.3** The majority of strong antiterminator interaction poses determined in multi-pose docking occurred in the bulge region.

Lastly, logP values (partition coefficient for octanol to water) values estimated in the Molecular Descriptors evaluation were used to determine which top-binding compounds could be ordered for use in the experimental screening series, as they need to be aqueous-soluble to yield reproducible results in the 5'-TAMRA-RNA experiments (unpublished results). Ten compounds were identified that were selective for one region of AM1A, were in the top 5% of Glide Emodel values (most negative) and had a logP value lower than 2.5 (Table 3.1). Of this set, acebutolol and palbociclib were ordered for use in experimental testing. In addition, amodiaquine, another compound in the MCE library, was also ordered as it had been previously identified as a compound with an exceptionally negative Glide Emodel value of -176.44, the strongest predicted bonding value in the multi pose compound docking, almost 20 units greater than the next strongest

Emodel value. However, amodiaquine was not initially considered for experimental screenings as its ligand poses bonded in more than one region in the multi-pose docking, either to helices A1 or A2. Ultimately, amodiaquine was additionally selected for screening due to its strong binding energy and to act as an indicator of the value of selecting consistently posed compounds. Amodiaquine and Acebutolol had their strongest binding energies when in poses interacting with A1, while palbociclib had the strongest binding interaction with the bulge (Figure 3.4). Intriguingly, all three of these compounds interacted with the antiterminator through electrostatic interactions of protonated amines and contained an aromatic moiety. These compounds were the first to be tested in the experimental fluorescence screening series.

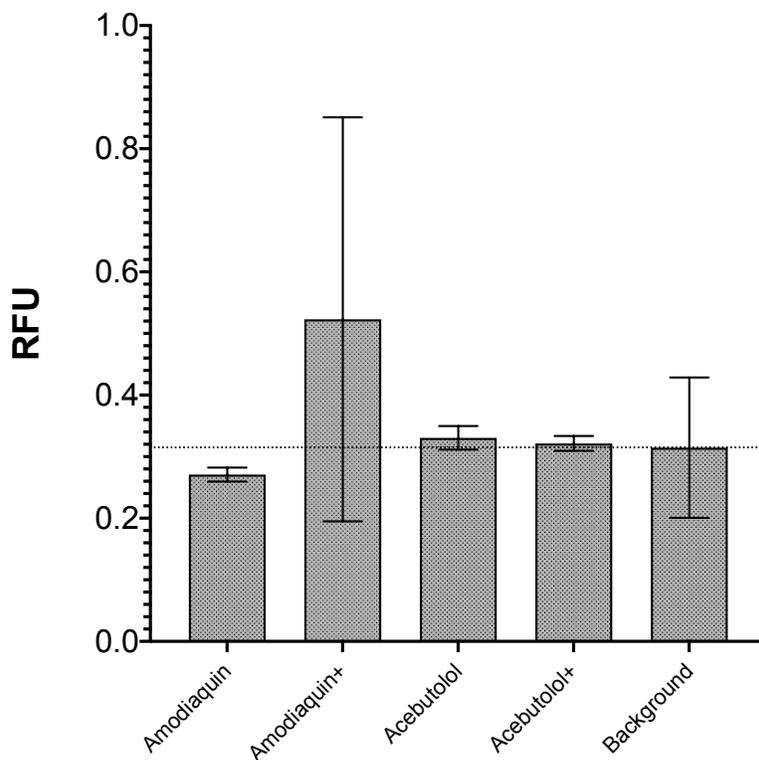
**Table 3.1** Ten compounds from the MCE library were identified to have favorable characteristics to be tested in experimental screenings.

Compound	Region	Glide Emodel	LogP
Avapritinib	A1	-116	1.9
Belotecan	A1	-112	1.6
Acebutolol	A1	-107	1.43
Tebipenem pivoxil	A1	-102	2.3
Palbociclib	Bulge	-150	1.8
Isoprenaline	Bulge	-107	1.4
Sotalol	Bulge	-115	0.2
DL-epinephrine	Bulge	-111	1.4
Almotriptan	A2	-95	1.6
Varencicline	A2	-74	0.8



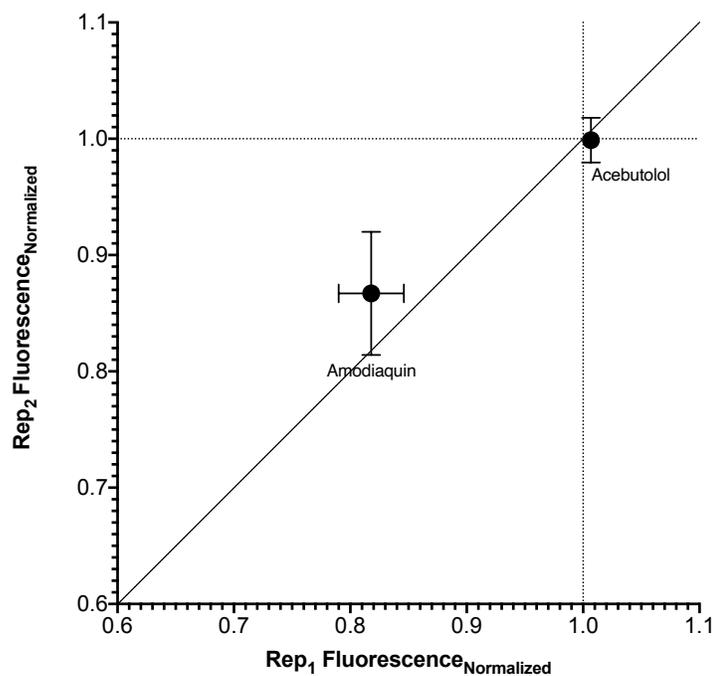
**Figure 3.4** The interactions of the top poses of acebutolol (A), palbociclib (B) and amodiaquine (C) with the antiterminator, as well as the array of poses for each compound are displayed.

3.2 Experimental screenings of FDA compounds indicate binding specificity in amodiaquine interactions with the antiterminator.



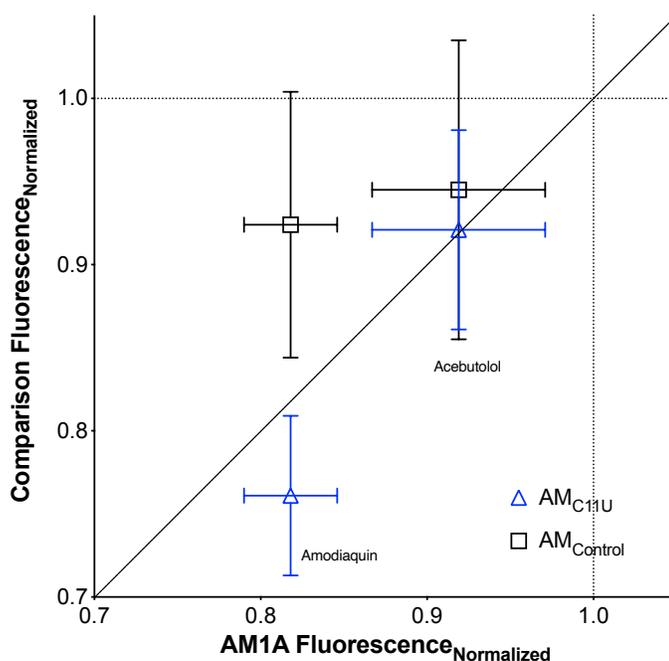
**Figure 3.5** Amodiaquine and acebutolol did not appear to have autofluorescence or RNA fluorescence effects. All compounds were tested in the reaction buffer, and + denotes mixtures with unlabeled AM1A.

Amodiaquine, acebutolol, and palbociclib were ordered and prepared for fluorescence experiments. Unfortunately, palbociclib had solubility issues at the concentration required for these experiments, 10  $\mu$ M, and as such it was not tested further. Acebutolol and amodiaquine were first tested against the reaction buffer and unlabeled AM1A in the control assay to test for disruptive fluorescence effects. Neither compounds showed greater fluorescence, measured with the arbitrary RFU value, than the background within error (Figure 3.5). As such, the compounds were then tested in the AM1A reproducibility assay.



**Figure 3.6** The reproducibility assay indicates consistent results for AM1A interaction with amodiaquine and acebutolol.

The AM1A reproducibility assay is a measure of the ligand ability and reproducible fluorescence values of the compounds interacting with 5'-TAMRA-AM1A. Acebutolol and amodiaquine produced consistent results for two averaged replicates, within error (Figure 3.6). However, the fluorescence modulation observed for the two compounds were distinct, with amodiaquine producing stronger deviation from fluorescence controls than acebutolol. This suggests that the amodiaquine has a stronger binding affinity to AM1A than acebutolol, or that acebutolol binds to AM1A in a manner that does not perturb the structural environment of the antiterminator, maintain a similar fluorescence to controls. With reproducible results determined for both compounds, amodiaquine and acebutolol were next screened against all three models of the antiterminator in the specificity assay.



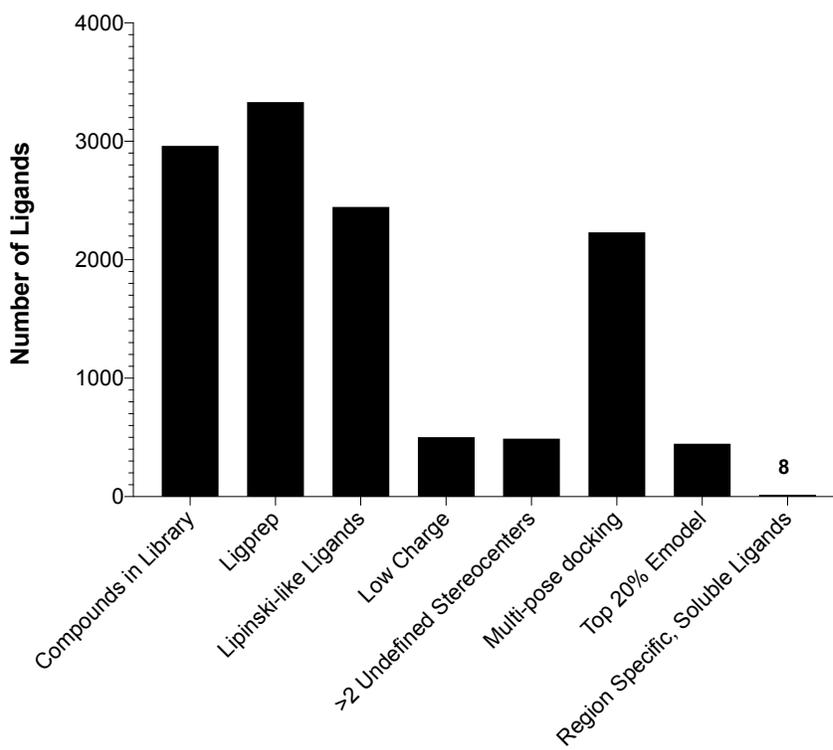
**Figure 3.7** The fluorescence specificity assay antiterminator model-specific modulation with amodiaquine, and nonspecific modulation with acebutolol.

As an indicator of antiterminator model ligand specificity, the fluorescence specificity assay tests each compound (10  $\mu$ M) against AM1A, AM<sub>C11U</sub>, and AM<sub>Control</sub> (100 nm) to determine differences in fluorescence modulation, reported as a normalized value. The fluorescence modulation for acebutolol was similar within error between all three antiterminator models (Figure 3.7). Acebutolol minorly affected the TAMRA fluorescence, with the greatest modulation observed in interaction with AM1A ( $0.919 \pm 0.052$ ). On the other hand, amodiaquine showed differences between each ligand model, with the greatest modulation observed in AM<sub>C11U</sub> ( $0.761 \pm 0.048$ ), followed by AM1A ( $0.818 \pm 0.028$ ), and with the least modulation observed in AM<sub>Control</sub> ( $0.924 \pm 0.080$ ). The lower fluorescence modulation in the bulge-free model AM<sub>Control</sub> is intriguing, as this model is the most sequentially and structurally distinct antiterminator model. While the immediate consideration is if the presence of the bulge is important for amodiaquine-

RNA interaction, it is also worth noting that the structure the helices in AM1A are more pinched than the usual A-form RNA, and as such the overall structure of AM1A and C11U are expected to be distinct from AM<sub>Control</sub>.

## 4. Results- ZINC Natural Metabolite Library

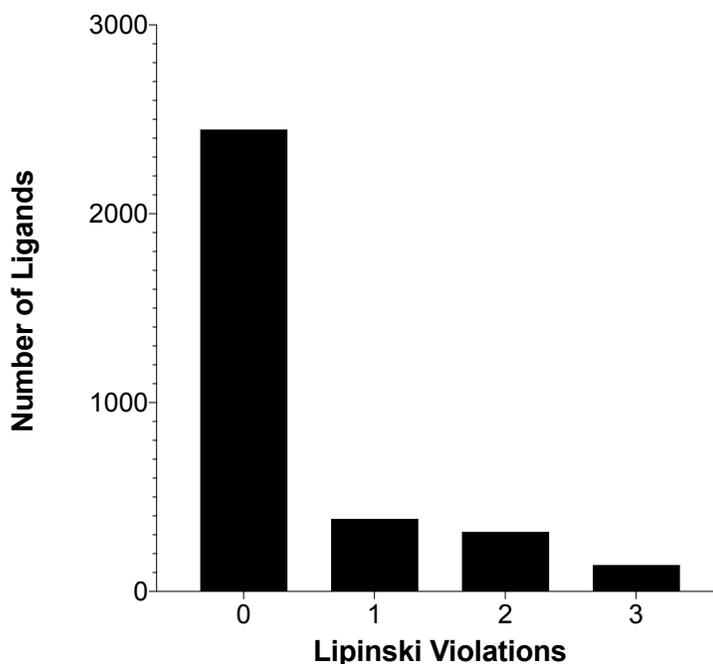
4.1 Computational preparation and docking of the ZINC library determined a subset of soluble, high affinity and positionally consistent compounds



**Figure 4.1** The computational process to identify strong ligand candidates to the T-box antiterminator identified eight ligands out of the ZINC natural metabolite library.

The ZINC (recursive acronym of ZINC is not commercial) database is an annotated repository for commercially available compounds, maintaining structural, chemical, and retailer information on compounds, as well as two-dimensional structural files in multiple formats. Within the ZINC database, a number of subsets of compounds have been curated, including the “natural-products” set of secondary metabolites. This subset was selected for investigation as a number of antibiotics have are either natural products or derivatives of natural products, such as aminoglycosides (Costales et al., 2020). The natural product library contains 48164 purchasable compounds, which would be computationally costly to generate ligands and docking poses for such a large library.

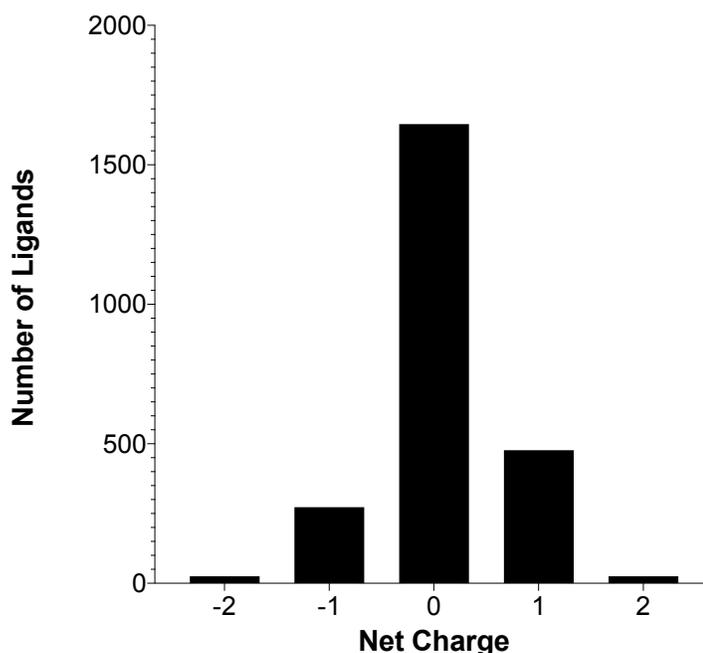
As such, the subset was further refined by filtering by the terms “for-sale” and “named”, reducing the library size to 2962 compounds, similar in scale to the MCE FDA-plus library. This refined ZINC natural product library was then screened in the hybrid assay workflow (Figure 4.1). The data file for the library compounds was collected and three-dimensional structures, stereoisomers, and tautomers were generated using the LigPrep program in Maestro with the OPLS\_2005 force field in pH conditions  $7.0 \pm 0.5$ . The LigPrep program yielded 3332 total potential ligands, with no dropped compounds.



**Figure 4.2** ZINC compounds mostly fell within the bounds of Lipinski’s rule of five, with no compounds violating all parameters.

Following the construction of the three-dimensional library of compounds in experimental conditions, the molecular descriptors for the compounds were collected using the Molecular Descriptors program in Maestro. A different refinement path was completed for this library as compared to the MCE library, but both started with the selection of druglike compounds. Compounds violating Lipinski’s rule of five were removed, reducing the library down to 2446 compounds, retaining the majority of the

library (Figure 4.2). It was observed that a number of compounds in this library had negative or no charge, which both of which are not favored in RNA-drug targeting due to electrostatic repulsion against the phosphate backbone and no salt-bridging to improve affinity, respectively (Figure 4.3) Compounds with charges of +1 and +2, a total of 502 structures, were selected for multi-pose docking. Lastly chiral compounds with more than 2 undefined stereocenters were removed to avoid the complications of replicating stereospecificity with compounds in the experimental assays, reducing the number of compounds down to 489 structures.



**Figure 4.3** Many of the druglike compounds identified in the ZINC library were filtered out as they did not contain favorable charges for RNA-binding.

A Glide docking of the selected compounds was completed recording up to 5 poses per ligand, yielding 2231 total poses. Interaction counts were again generated for these poses using the Interaction Fingerprint program, and the total interactions for each antiterminator region were again used to sort poses by the highest interaction totals for each region. Compounds with the most negative Emodel value, indicating the best

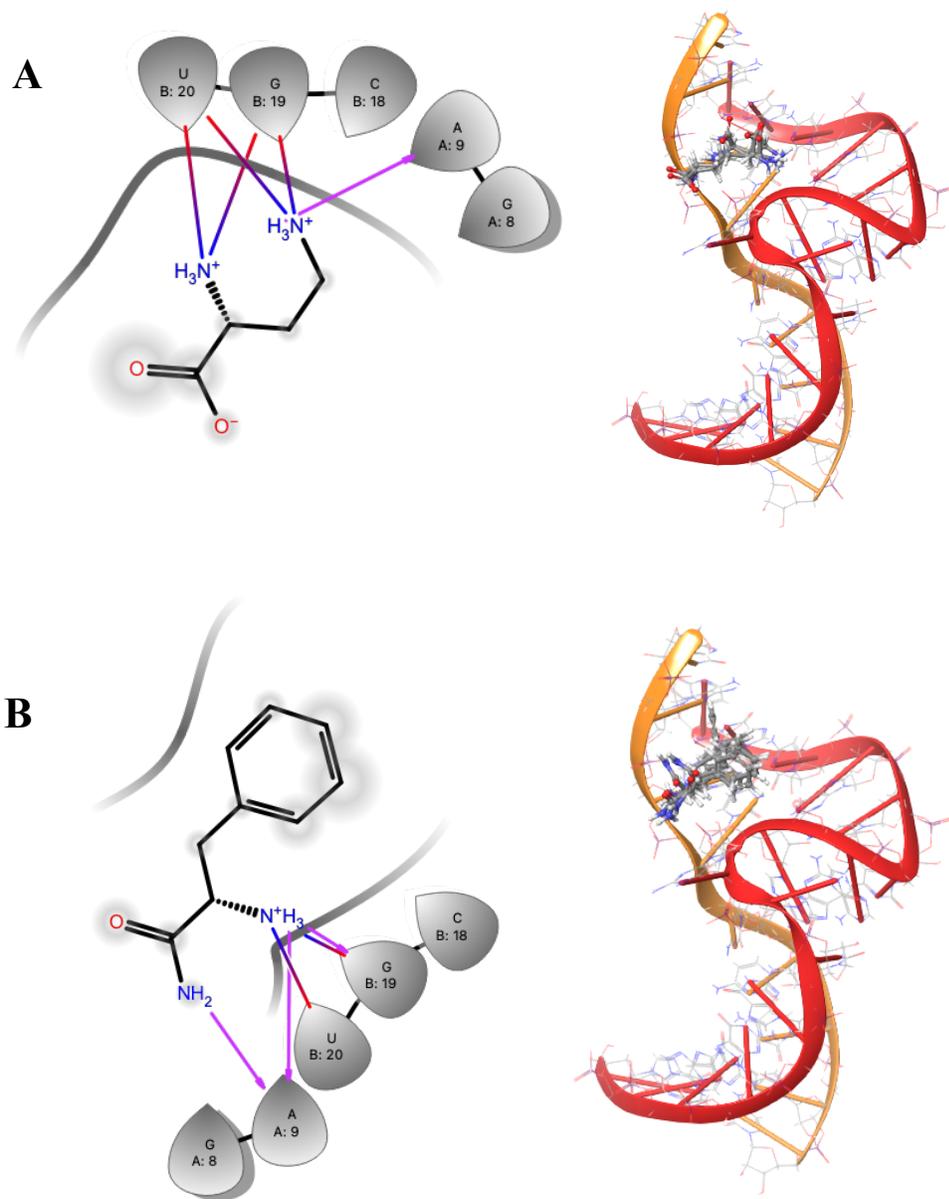
interaction energy, of binding to the antiterminator for each compound, were then evaluated for consistent poses.

**Table 4.1** Eight compounds from the ZINC library were identified to have favorable characteristics to be tested in experimental screenings.

Compound	Region	Glide Emodel	QPLogS
Chimonanthin	A1 and Bulge	-130	0.36
Spectinomycin	Bulge and A2	-125	0.07
Oxilofrine	Bulge and A2	-110	-0.47
2,4-Diaminobutyrate	Bulge and A2	-104	1.74
L-phenylalaninamide	Bulge and A2	-96.6	1.32
Phenylpropanolamine	Bulge	-94.7	-0.13
Desglymodrine	Bulge	-93.9	-0.47
Phenylpropanolamine	Bulge	-93.6	-0.09

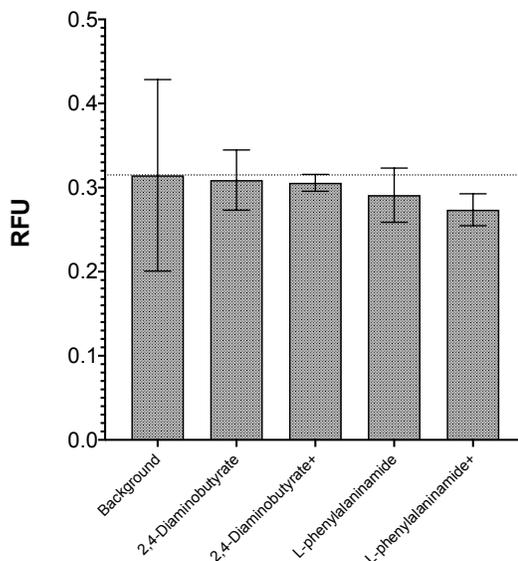
The compounds were sorted by the number of interactions to each region, and the top 20% of compounds for each region were annotated with the top region. The now annotated ligands were now sorted by Emodel, and the top 20% of Emodel compounds were filtered to remove compounds with strong interactions in multiple regions. Compounds were also included that did not preferentially bind to one region, but consistently interacted with the antiterminator across two regions. Nonduplicate compounds were finally selected by their solubility, this time determined by a more direct value of solubility, the Molecular Descriptors predicted solubility, QPlogS. Eight compounds were identified that were selective for one region of AM1A, were in the top 20% of Glide Emodel values (most negative) and had a QPlogS value greater than -0.5 (Table 4.1). Of this set, 2,4-diaminobutyrate and L-phenylalaninamide were ordered for use in experimental testing. These compounds both bound to the A2 and bulge regions

through hydrogen bond or salt bridges between their two amines and the backbone of AM1A (Figure 4.4).



**Figure 4.4** The interactions of the top poses of 2,4-diaminobutyrate (A), and L-phenylalaninamide (B) with the antiterminator, as well as the array of poses for each compound are displayed.

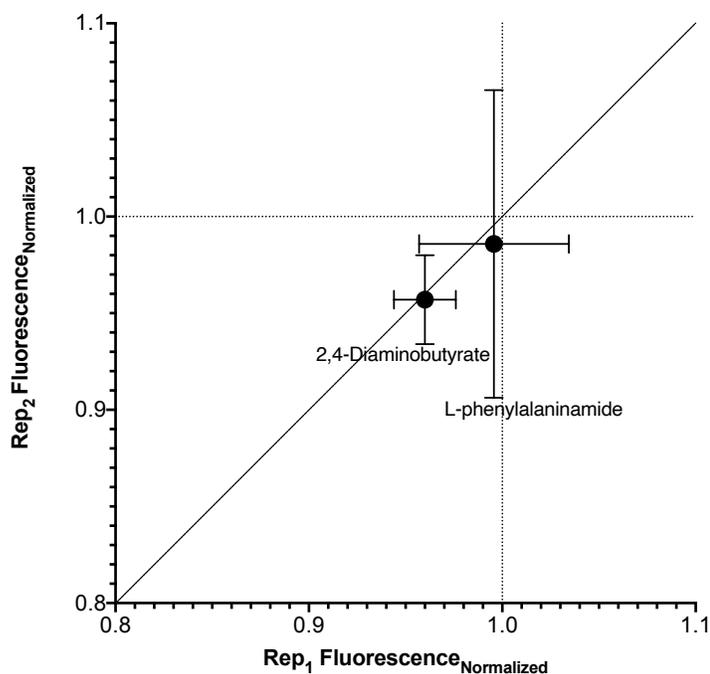
4.2 Experimental screenings of ZINC compounds indicated nonspecific interactions with the antiterminator.



**Figure 4.5** 2,4-diaminobutyrate and L-phenylalaninamide did not appear to have autofluorescence or RNA fluorescence effects. All compounds were tested in the reaction buffer, and + denotes mixtures with unlabeled AM1A.

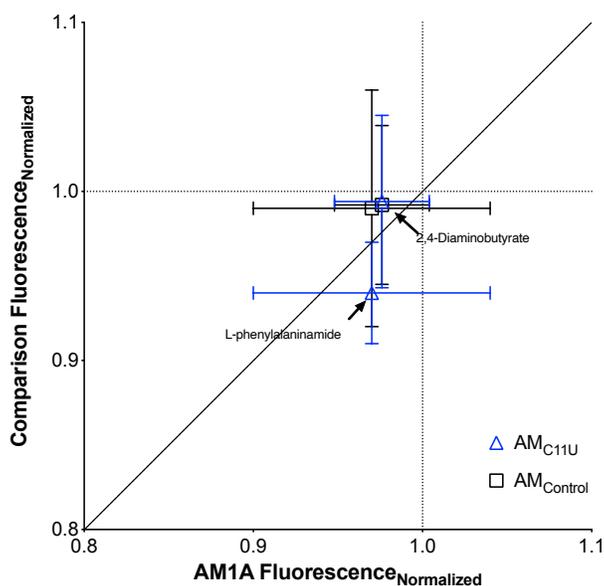
L-phenylalaninamide and 2,4-diaminobutyrate were ordered and prepared for fluorescence experiments. The compounds were tested in the fluorescence control assay, and neither compound displayed greater fluorescence increases beyond the background well fluorescence within error (Figure 4.5). As such, the compounds were then tested in the AM1A reproducibility assay.

In the AM1A reproducibility assay, 2,4-diaminobutyrate appeared to have a reproducible minor fluorescence modulation of AM1A, while L-phenylalaninamide showed a lesser modulation, albeit with a larger error (Figure 4.6) The results of 2,4-diaminobutyrate mixed with AM1A does deviate from the background fluorescence, suggesting that there is some interaction or modulatory effect occurring between it and AM1A. On the hand, L-phenylalaninamide displayed no fluorescence modulation within error with AM1A, meaning it may have a null effect on AM1A structure or does not interact with the antiterminator. Both compounds were subsequently tested in the fluorescence specificity assay.



**Figure 4.6** The reproducibility assay indicates consistent results for AM1A interaction with 2,4-diaminobutyrate and L-phenylalaninamide

2,4-diaminobutyrate and L-phenylalaninamide were screened against 5' TAMRA-labelled AM1A, AM<sub>C11U</sub>, and AM<sub>Control</sub> to determine differences in fluorescence modulation. Both compounds exhibited normalized fluorescence values similar within error between all three antiterminator models (Figure 4.7). Both compounds only display minor modulation in TAMRA-RNA fluorescence compared to controls for each model, with the greatest modulation observed in interaction with AM1A for both compounds ( $0.976 \pm 0.028$  for 2,4-diaminobutyrate,  $0.970 \pm 0.070$  for L-phenylalaninamide). As both compounds were predicted in computational simulations to interact with the bulge, it is interesting that there was no observed fluorescence specificity between the bulge free model AM<sub>Control</sub> and the bulge containing models in either compound.

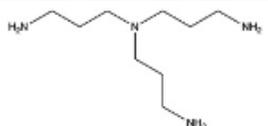
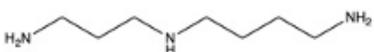
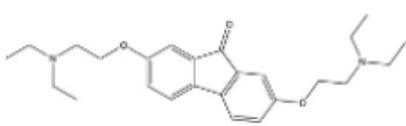
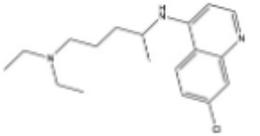
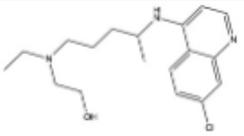
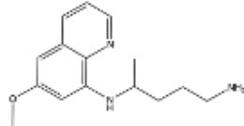


**Figure 4.7** The fluorescence specificity assay showed nonspecific effects for both 2,4-diaminobutyrate and L-phenylalaninamide.

## 5. Results – Lab Compound Training Set

In addition to compounds selected from computational library screening a number of other compounds were also tested using the terminal target labelling assays. These compounds were then computationally docked to the antiterminator structure, allowing investigation into how reversing the project workflow might affect results. Broadly, these additional compounds can be separated into categories, polycyclic heterocycles and aliphatic polyamines (Table 5.1).

**Table 5.1** Seven additional compounds were tested in the experimental screenings, categorized into polyamines and polycyclic heterocycles.

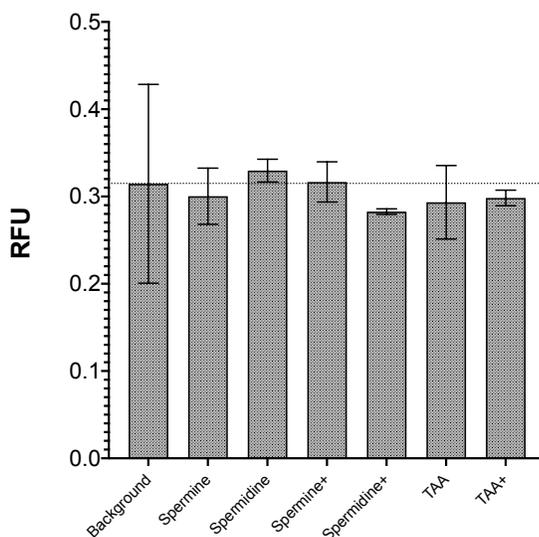
Compound	Type	Structure	MW (g/mol)
Tris(3-aminopropyl)amine	Polyamine		188.31
Spermidine	Polyamine		145.25
Spermine	Polyamine		202.34
Tilorone	Polycyclic Heterocycle		483.47
Chloroquine	Polycyclic Heterocycle		319.87
Hydroxychloroquine	Polycyclic Heterocycle		335.87
Primaquine	Polycyclic Heterocycle		259.35

The nitrogenous heterocycles were previously identified as potential candidates for another RNA target investigated in the Hines Lab (unpublished results) and were

tested in this project to see if their structure would be favorable for binding another RNA structure. As well, the model-specific results observed in screening amodiaquine encouraged screening these compounds, as many of them also have a quinoline ring. Quinoline, a nitrogenous heteroatomic cycle, has a number of derivative compounds used in therapeutic treatments. In fact, quinoline derivatives have been used in the treatment of malaria, rheumatic diseases, lupus, and as an anticancer treatment (Al-Bari, 2015). Some 9-aminoquinoline compounds are also known to exert antiviral effects, including inhibition of HIV replication (Savarino et al., 2003). The diverse pharmacological effects of quinoline compounds, including decreased DNA, RNA, and protein expression (Al-Barie, 2015) could implicate quinoline derivatives as having promiscuous binding activity in cellular environments. Polyamines have previously been identified as being able to represent an unknown cofactor in *in vitro* transcription assays using the T-box antiterminator (Liu et al. 2016). Investigating both sets of these compounds provided insight into using the assay to identify structure-activity relationships.

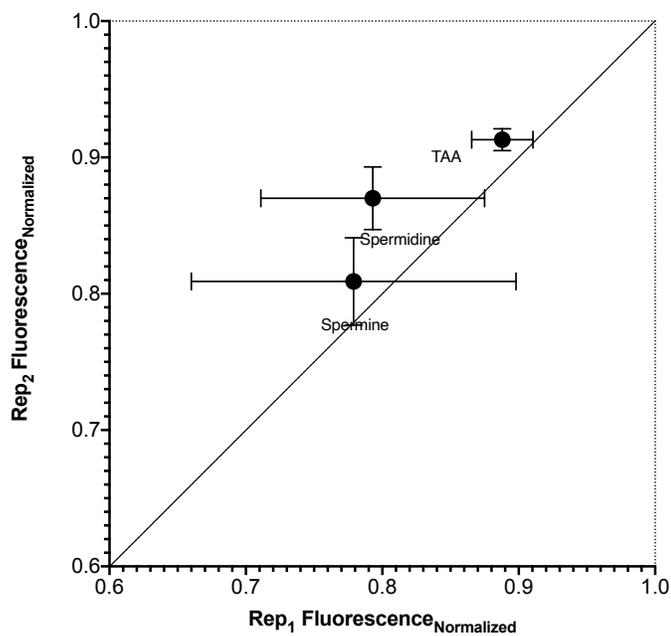
### *5.1 Experimental screenings of polyamines displayed antiterminator model-specific fluorescence modulation.*

Spermidine and spermine are known to bind the antiterminator, and Tris(3-aminopropyl)amine (abbreviated to TAA) was identified in a previous literature search of polyamines and docked to the antiterminator in a computational study. The compounds were tested in the fluorescence control assay, and no fluorescence increases beyond the background well signal was observed within error (Figure 5.1). The polyamines were then screened for a reproducible signal with AM1A.



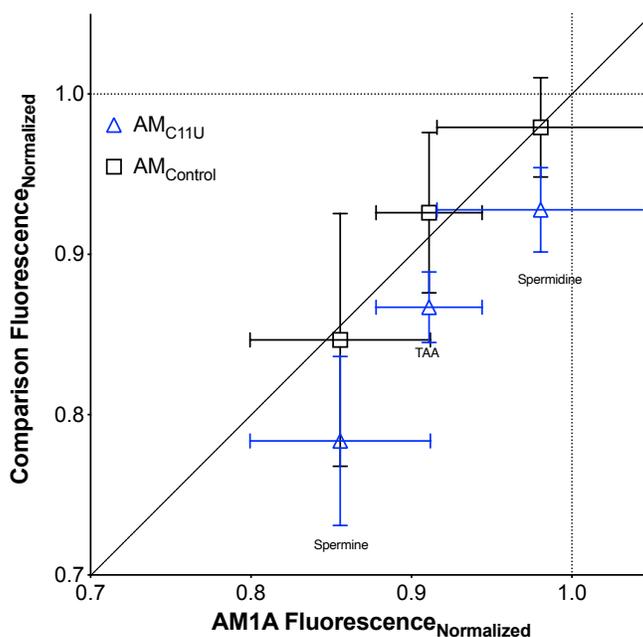
**Figure 5.1.** The polyamines did not display autofluorescence effects. All compounds were tested in the reaction buffer, and + denotes mixtures with unlabeled AM1A.

In the AM1A reproducibility assay, all three compounds were reproducible within error (Figure 5.2). While the larger error present in the first replicate makes interpretation dubious, it appears that spermine has the greatest modulation on AM1A structure, followed by spermidine then TAA. With acceptable values in the reproducibility screening, the polyamines were then tested in the specificity screening.



**Figure 5.2** The reproducibility assay indicates consistent results for the polyamines.

While spermidine showed no differences within error between AM models, spermine and TAA displayed differences in fluorescence modulation between the AM<sub>C11U</sub> and AM1A, while no differences were displayed within error between AM1A and AM<sub>Control</sub> (Figure 5.3). An interesting trend among the polyamines is that the greatest fluorescence modulation in all antiterminator models was observed in spermine, followed by TAA, and finally in spermidine. This trend correlates to the expected charge state of these compounds in the assay conditions, as spermine is expected to have a +4 charge, TAA is able to have both +3 and +4 states, and spermidine has a +3 charge at pH = 6.5.

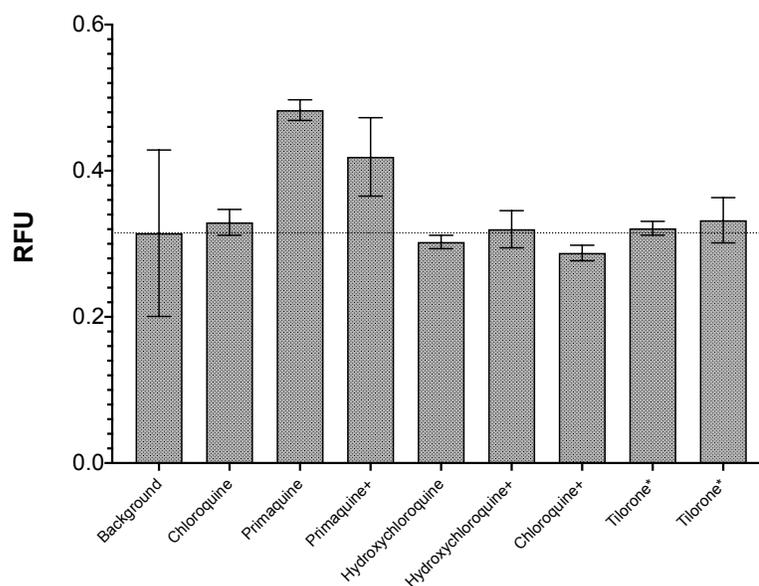


**Figure 5.3** The fluorescence specificity assay indicates specific effects between AM1A and AM<sub>C11U</sub> for TAA and spermine

### 5.2 Experimental screenings of polycyclic heterocycles identified antiterminator model-specific fluorescence in all compounds

In the fluorescence control assay, and no fluorescence increases beyond the background well signal was observed within error except for primaquine in the reaction buffer (Figure 5.4). However, as the mixture of AM1A with primaquine fell within error

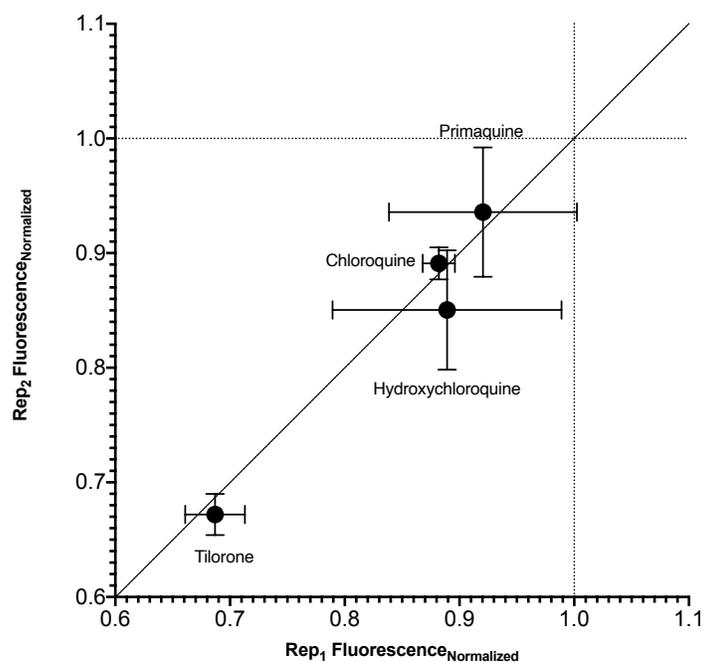
of background fluorescence, primaquine was included with the rest of the compounds when screened for a reproducible signal with AM1A. In the AM1A reproducibility assay, all compounds displayed reproducible signals within error (Figure 5.5). The compounds containing a quinoline ring - chloroquine, hydroxychloroquine, and primaquine- all had lower modulation values than tilorone.



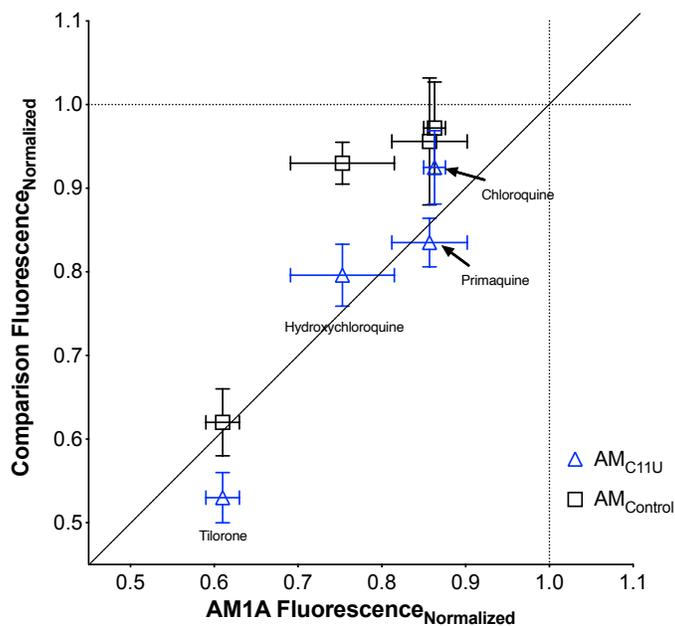
**Figure 5.4** The polycyclic heterocycle compounds were screened in the fluorescence control assay. All compounds were tested in the reaction buffer, and + denotes mixtures with unlabeled AM1A. \* denotes that a small amount of hydrochloric acid (0.29 mM) was added to the controls for tilorone to match solubility conditions for the compound.

The heterocyclic compounds were screened in the specificity assay (Figure 5.6).

Hydroxychloroquine and primaquine showed similar modulation between AM1A and  $AM_{C11U}$ , and a lower modulation in  $AM_{Control}$ . Chloroquine however, showed greater fluorescence modulation in AM1A than the other two models. Tilorone displayed greater overall fluorescence modulation in all antiterminator models than the quinoline compounds, with similar values in AM1A and  $AM_{Control}$  and greater fluorescence modulation in  $AM_{C11U}$ .



**Figure 5.5** The reproducibility assay indicates consistent results for the polyamines.



**Figure 5.6** The fluorescence specificity assay indicates specific effects between AM1A and AM<sub>C11U</sub> for TAA and spermine

5.5 Training set compounds were computationally docked to AM1A.

Following completion of experimental screening of the training set compounds their structure data files were collected from the PubChem database and three-

dimensional structures of the compounds were generated using the LigPrep program in Maestro. The attributes of the compounds were detailed using the Molecular Descriptors, and the compounds were then docked to the antiterminator using the Glide program in Maestro. The output file from Glide docking was then processed in the Interaction Fingerprint program to identify the ligand-RNA interactions by residue and classify the compounds by the top binding region (Table 5.2). The polyamines displayed the best Emodel values and preferentially bound the bulge and helix A2. The polycyclic heterocycles had less negative, weaker Emodel values, and had greater variety in region binding.

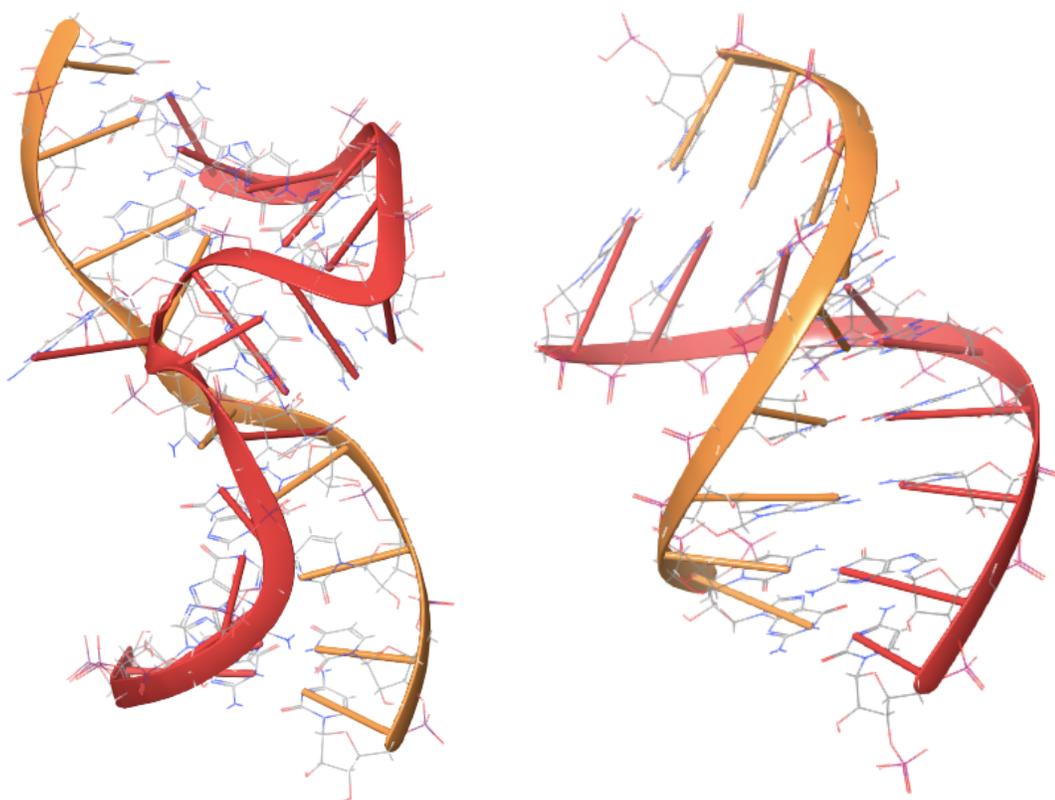
**Table 5.2** Seven additional compounds were tested in the experimental screenings, categorized into polyamines and polycyclic heterocycles.

<b>Compound</b>	<b>Charge</b>	<b>Top binding Region</b>	<b>Top Emodel Value</b>
Tris(3-aminopropyl)amine	3,4	Bulge, A2	-259
Spermidine	3	Bulge	-186
Spermine	4	Bulge, A2	-260
Tilorone	2	A2	-138
Chloroquine	2	Bulge	-116
Hydroxychloroquine	2	A1	-161
Primaquine	1	A2	-90.4

## 6 Results comparison of computational studies to fluorescence results

### 6.1 Development and docking for a computational $AM_{Control}$ model

To develop a larger vista of how docking scores translate to fluorescence binding modulation in the antiterminator studies, a model of  $AM_{Control}$  was developed in Maestro (Figure 6.1). As a simple A-form helix, the model was developed using the Builder tool in Maestro by sequence, excluding the UUCG tetraloop to match the AM1A NMR solution structure. The RNA was then corrected to include the G-U wobble pair at the sixth position in PyMol, and lastly a grid was developed for molecular docking to the sequence, with similar parameters to the AM1A grid.



**Figure 6.1** Computational structures of AM1A (left) and  $AM_{Control}$  (right) were prepared using Protein Prep Wizard and Grid Generation in Maestro.

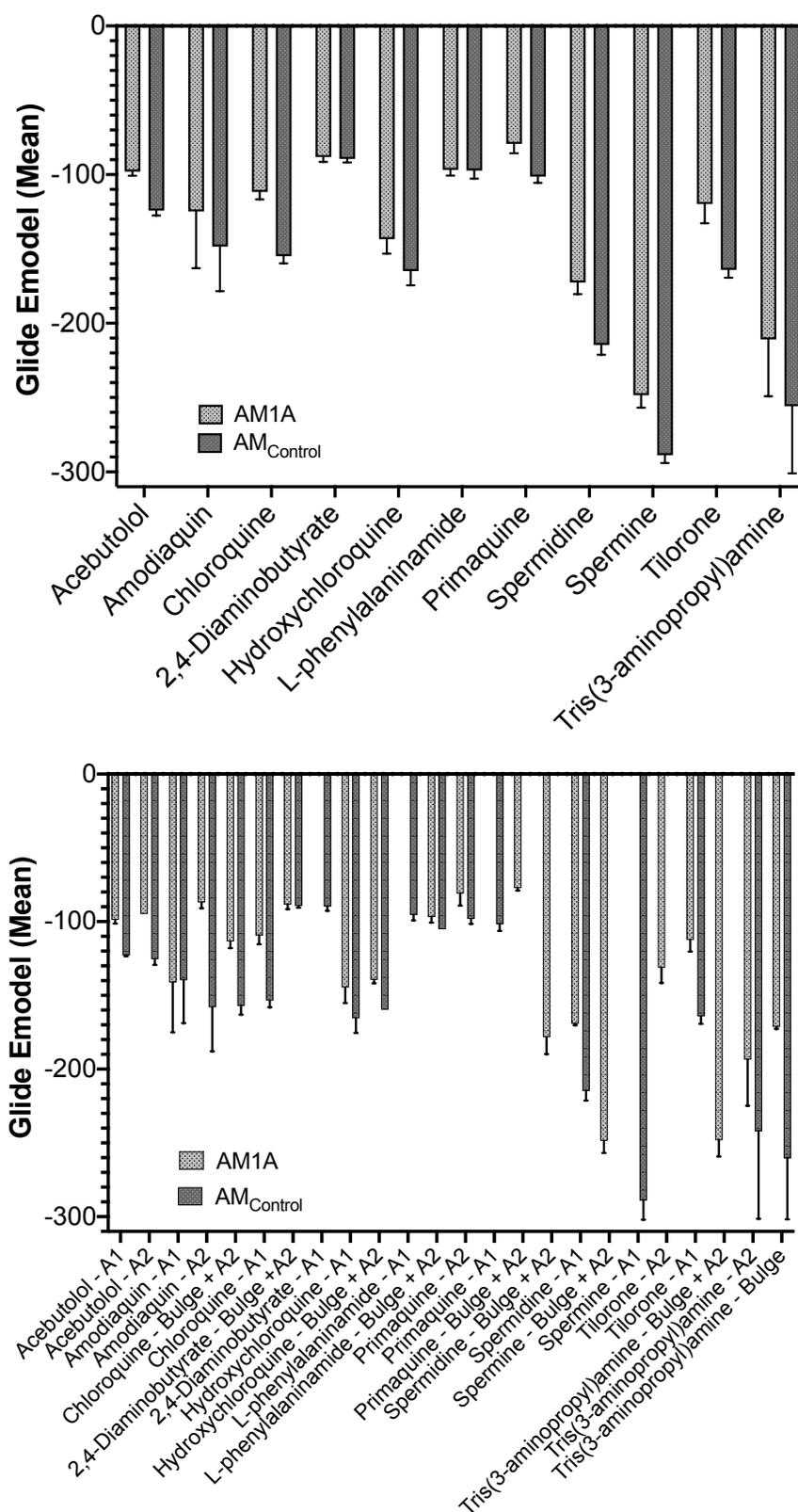
Following the generation of its grid, all compounds previously tested in the fluorescence experiments were docked to  $AM_{Control}$ . The predicted binding compound

energies were then compared between the two antiterminator models. Generally, the Emodel values determined for poses to AM<sub>Control</sub> were more negative than the compounds pose docked to AM1A (Figure 6.2).

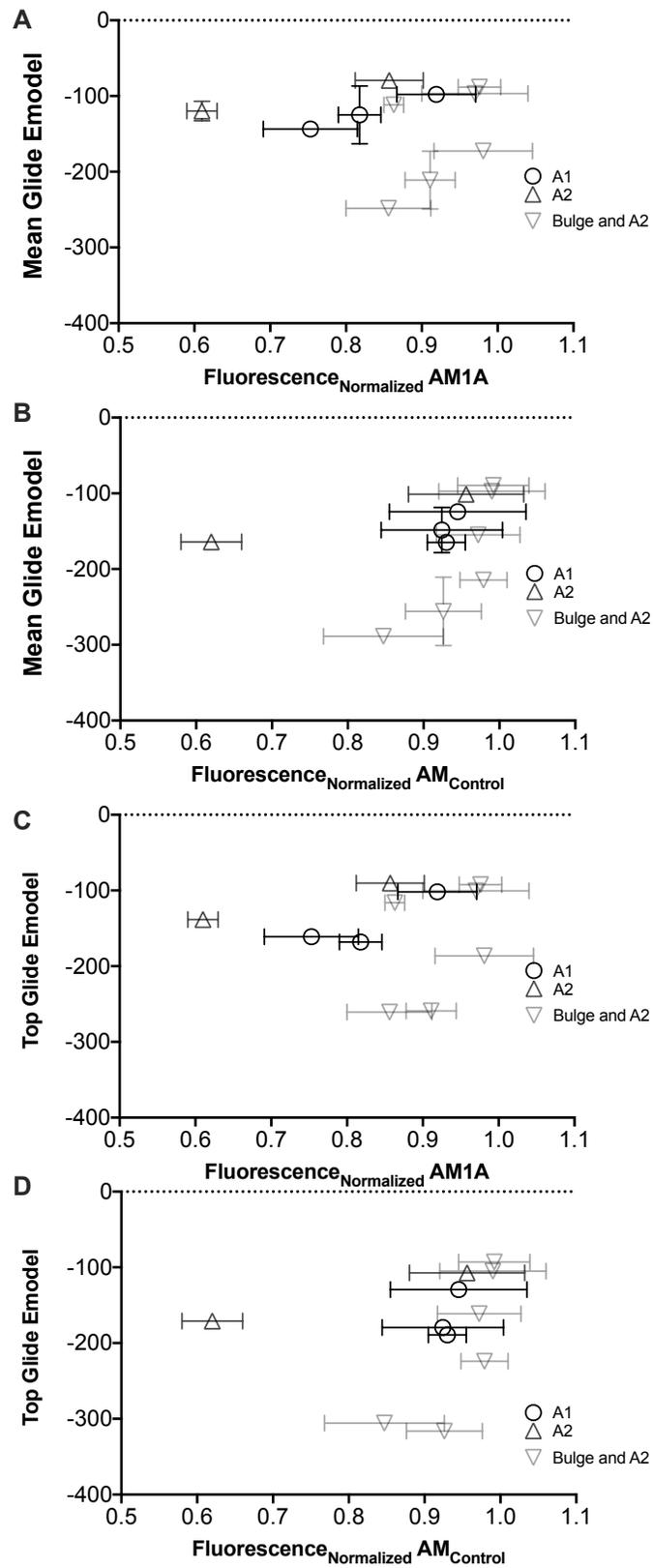
### *6.2 Comparison of computational and Emodel values reveals trends in predictive value of computational docking by region.*

Comparison of normalized fluorescence scores for compound interaction with AM1A and AM<sub>Control</sub> to Glide Emodel values for docking poses of the compounds to the models of the antiterminator sequences reveal trends in fluorescence modulation by predictive region, as identified by the sum interaction fingerprints for each region in AM1A (Figure 6.3). While the data sets are small for A2 and A1 selective compounds, it appears the moderate differences in Emodel values for ligands binding these regions lead to greater differences in fluorescence modulation, as these compounds displayed greater shifts in fluorescence modulation over slight changes in Emodel value. The A2 binding compounds showed the greatest sensitivity to glide Emodel scoring, as difference in approximately 40 units for this value lead to a difference of .25 in normalized fluorescence values when tested against AM1A, and a difference of approximately 60 units in the Emodel of binding to AM<sub>Control</sub> lead to a difference of 0.3 in normalized fluorescence. The A1 binding compounds were less sensitive than A2 compounds, but still generally displayed increased fluorescence modulation as the Emodel value calculated for compounds binding this region decreased. The bulge + A2 region binding compounds appeared the least sensitive to predicted binding Emodel values, as a comparison between the two “endpoints” of these compounds revealed that a difference in nearly 160 Emodel units only produced a difference of 0.12 in normalized fluorescence

values for AM1A, and a difference of approximately 210 Emodel units only produced a difference of 0.15 in normalized fluorescence between the same compounds in AM<sub>Control</sub>. In comparison of the two compound models, it is evident that compounds that bound to A2 region displayed similar trends in AM1A and AM<sub>Control</sub>, while A1 compounds were slightly less sensitive in when interacting with AM<sub>Control</sub>, and bulge/A2 models, with the bulge representing the G-U wobble pair only in AM<sub>Control</sub>, were slightly more sensitive than when interacting with AM1A. It is interesting that the structural differences between the two antiterminator models resulted in different correlations between region binding compound Emodel interaction values, and it is encouraging that in both cases, stronger Emodel values, which are predictive computational values of interactions, corresponded to stronger fluorescence modulation values when compounds were evaluated by predicted region of interaction. It is also worth noting that the same trends were noted in using both the mean values of Glide Emodel (6.3A and 6.3B) and the most negative, optimal Glide Emodel values (6.3C and 6.3D), suggesting these top scores may have similar predictive value to multi-pose docking.



**Figure 6.2** Compounds were docked to both AM1A and AM<sub>Control</sub>. Interactions with AM<sub>Control</sub> had greater mean Emodel values than compounds docked to AM1A for all compounds (Top) and were generally higher by region as well (bottom). Error bars are standard deviation for all ligand pose Emodel values.



**Figure 6.3** Comparison of normalized fluorescence in AM<sub>1A</sub> and AM<sub>Control</sub> reveals predictive value by region for both mean Emodel (A and B) and Top Emodel values (C and D).

## 7. Conclusion

**Figure 7.1** Compilation of results determined in the combined assay.

Compound	Interaction Fingerprint Region	Normalized Fluorescence (AM1A)	Normalized Fluorescence (AMC11U)	Normalized Fluorescence (AM <sub>Control</sub> )	Specificity	Glide Emodel (AM1A)	Glide Emodel (AM <sub>Control</sub> )
Acebutolol	A1	0.918±0.051	0.921±0.060	0.944±0.090	AM1A ≈ AMC11U AM1A ≈ AM <sub>Control</sub> AMC11U ≈ AM <sub>Control</sub>	-102	-129
Amodiaquimodne	A1, A2	0.818±0.028	0.760±0.048	0.924±0.080	AM1A < AMC11U AM1A > AM <sub>Control</sub> AMC11U > AM <sub>Control</sub>	-168	-179
Chloroquine	A2, Bulge	0.863±0.013	0.925±0.043	0.972±0.054	AM1A < AMC11U AM1A < AM <sub>Control</sub> AMC11U ≈ AM <sub>Control</sub>	-116	-161
2,4-Diaminobutyrate	A2, Bulge	0.975±0.028	0.994±0.050	0.992±0.047	AM1A ≈ AMC11U AM1A ≈ AM <sub>Control</sub> AMC11U ≈ AM <sub>Control</sub>	-92.4	-92.9
Hydroxychloroquine	A1	0.752±0.061	0.796±0.036	0.929±0.024	AM1A ≈ AMC11U AM1A > AM <sub>Control</sub> AMC11U > AM <sub>Control</sub>	-161	-189
L-phenylalaninamide	A2, Bulge	0.97±0.07	0.94±0.03	0.99±0.07	AM1A ≈ AMC11U AM1A ≈ AM <sub>Control</sub> AMC11U ≈ AM <sub>Control</sub>	-100	-105
Primaquine	A2	0.856±0.045	0.834±0.029	0.956±0.075	AM1A ≈ AMC11U AM1A > AM <sub>Control</sub> AMC11U > AM <sub>Control</sub>	-90.4	-107
Spermidine	A2, Bulge	0.980±0.064	0.927±0.026	0.979±0.031	AM1A < AMC11U AM1A ≈ AM <sub>Control</sub> AMC11U > AM <sub>Control</sub>	-186	-224
Spermine	A2, Bulge	0.855±0.056	0.783±0.052	0.846±0.078	AM1A < AMC11U AM1A ≈ AM <sub>Control</sub> AMC11U > AM <sub>Control</sub>	-260	-306
Tilorone	A2	0.61±0.02	0.53±0.03	0.62±0.04	AM1A < AMC11U AM1A ≈ AM <sub>Control</sub> AMC11U > AM <sub>Control</sub>	-138	-171
Tris(3-aminopropyl)amine	A2, Bulge	0.911±0.033	0.866±0.021	0.926±0.050	AM1A < AMC11U AM1A ≈ AM <sub>Control</sub> AMC11U > AM <sub>Control</sub>	-259	-316

The conjunction of computational and experimental work on the antiterminator is an exciting piece of the puzzle in research on this RNA and looks to have potential to yield effective predictions on binding, and hopefully, eventual drug capability. The data and analysis in this thesis worked to combine the methods of previous research to refine the approach to primary screenings of candidate compounds that may bind the T-box antiterminator. The anticipated outcome of this thesis project was the optimization of an iterative primary high-moderate throughput screening for antiterminator drug candidate compounds. The implementation of this process also provided a small library of compounds which have variable affinities and specificities to the models of the T-box antiterminator.

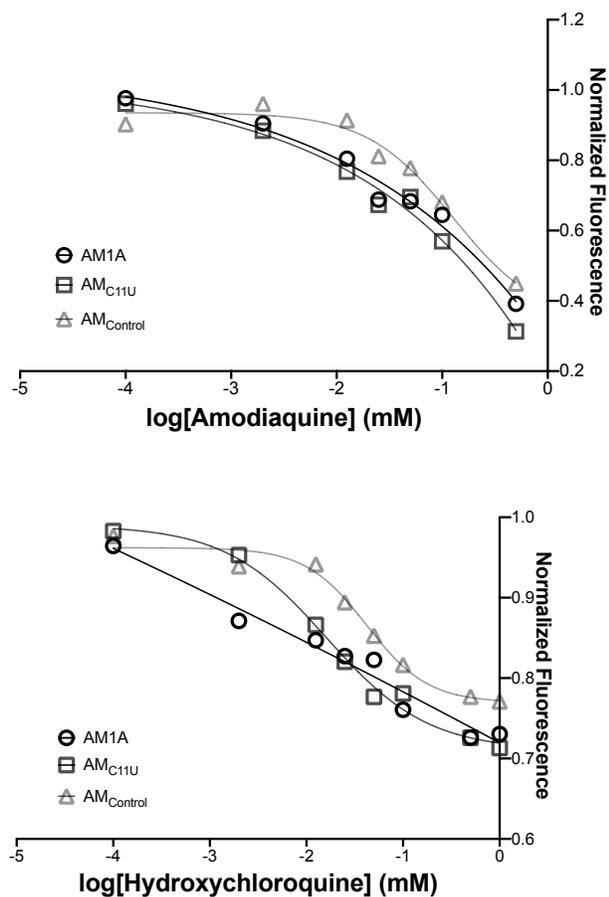
While no concrete relationship between experimental binding specificity and computational modeling was determined, a few trends were noted that relate the two processes together. It appears as though the binding regions, composed of important RNA secondary structural motifs within the antiterminator, to which compounds bind in computational docking displayed a correlation between Glide Emodel and degree of fluorescence modulation in the 5'-TAMRA labelled antiterminator models. If this trend is scalable to larger libraries, this would allow for an initial indication of relative strength between compounds that bind to the same region of the T-box riboswitch. This study also indicated that the strongest single ligand poses for a compound, as determined by the most negative Glide Emodel value, displays a similar correlation as the mean Glide Emodel value with fluorescence modulation in the experimental screenings. This could mean that larger libraries could be screened at a lower storage cost by collecting a single pose or fewer poses for determining candidate compounds, and that compounds that

bound more than on region in the computational docking may need to be reconsidered as candidate compounds.

In addition to the trends described above, this project identified eight out of the eleven compounds tested as having some model-specific effects as determined in the fluorescence specificity screening. All of these compounds have at least one amine present which participated in hydrogen bonding or electrostatic interaction with the antiterminator and can be grouped into two categories of compounds: polyamines and polycyclic heterocycles. The polyamines spermine, spermidine, and Tris(3-aminopropyl)amine showed similar specificities, having greater effects on AM<sub>C11U</sub> than the other two models. The polycyclic heterocycles had more heterogeneity in their specific effects on fluorescence in the dye-bound models, though in general the quinoline compounds had a stronger effect on AM<sub>C11U</sub> and AM<sub>1A</sub> than AM<sub>Control</sub>, while tilorone had a greater effect on AM<sub>C11U</sub> than the other two models, similar to the polyamines. Polycyclic heterocycles have previously been implicated as valuable scaffolds for small-molecule RNA-binding compounds (Juru et al., 2021), which has been supported in the results of this project. Additionally, these results suggest quinoline rings may be another scaffold that should be considered in structure-activity studies for RNA-binding compounds. As previously discussed, RNA-binding ligands often contain nitrogens, aromatic rings, and a positive charge. These physicochemical traits paired with the multiple physiological effects produced by quinoline derivatives suggest it may have a general affinity for RNA, which could be exploited in future drug development schemes.

The results of this project should be further corroborated with more intensive, moderate to low-throughput studies which can more definitively identify binding

differences between compounds and antiterminator models. Dose response studies, fluorescence anisotropy assays, and in vitro transcription assays have all been previously used in the Hines lab to identify ligand effects on compounds, though dose responses assay would be the most efficient to implement to study multiple antiterminator model RNAs. Such studies can be used to determine  $K_d$  values and sigmoidal response curves to provide a definitive comparison different RNA-ligand binding affinity. To this end, following completion of computational and fluorescence studies of amodiaquine and hydroxychloroquine they were subsequently tested against the TAMRA-antiterminator RNAs in a dose-response screening, testing across multiple magnitudes of concentration (Figure 7.1). These preliminary results appear promising as the different antiterminator models display different curve responses to compound concentrations. These studies will need to be completed with a greater concentration range and with more intermediate concentrations to better define the sigmoidal curve and plateaus, as well as to determine the reproducibility of results.



**Figure 7.1** Initial dose response studies of amodiaquine and hydroxychloroquine with TAMRA-antiterminator models displayed fluorescence differences consistent with single-concentration screenings.

The computational filtering and analysis developed in this study could also be expanded. A homology model of AM<sub>C11U</sub> using the solution structure of AM1A as a basis model could be developed such that each antiterminator model could be represented *in silico* as well as in experimental methods for comparison between models and between computational and experimental results. As previously stated, results from the compounds tested in fluorescence experiments in this project indicate that the strongest pose was as strong of a predictor of fluorescence modulation as the mean of all poses, opening the door for larger libraries to be tested as well as reconsideration of compounds with strong Glide Emodel values and inconsistent poses identified in the MCE and ZINC

libraries tested. Additional libraries should as be tested, such as the soluble fragments library from MedChemExpress or the  $\log P < -1$  tranche from ZINC. While specificity is not well determined by computational docking and analysis to the T-box antiterminator, this project has shown that annotation of important regions and compound-RNA interactions in computational studies can be used to indicate relative strengths of compound binding. This can assist in efficient ranking of compound potential for experimental studies, aiding in antiterminator drug discovery research.

## 8. Methods

### 8.1 RNA preparation

5'-TAMRA labelled model RNAs were ordered from Dharmacon Inc (Colorado). The RNAs were dissolved in distilled, deionized water and dialyzed against a dialysis buffer: 10 mM MOPS, 0.01 mM EDTA, pH=6.5). The RNAs were dialyzed against a liter of the buffer for 8 hours, followed by a second liter of buffer for an additional 16 hours. The concentration of each RNA was measured using absorbance at 260 nm and calculated with the extinction coefficient provided from Dharmacon. The purified RNA stocks were wrapped in aluminum foil and stored at -20 °C.

RNA	Sequence	Extinction coefficient (M <sup>-1</sup> cm <sup>-1</sup> )	Concentration of Stock (μM)
AM1A	TAM- GAGGGUGGAACCGCGCUUCGGCGUCCCUC	298680	112.9
AMC11U	TAM- GAGGGUGGAAUCGCGCUUCGGCGUCCCUC	301980	93.6
AM <sub>Control</sub>	TAM-GAGGGGCGCUUCGGCGUCCCUC	229980	106

### 8.2 Compound Preparation

Compounds tested in fluorescence assays were ordered from a number of companies. Upon arrival, the compounds were stored in conditions recommended in their SDS until they were diluted. Compounds were diluted in distilled, deionized water to a

concentration below their maximum solubility, between 1 and 50  $\mu\text{M}$ . Following dilution, compounds were stored at  $-20\text{ }^{\circ}\text{C}$ .

### *8.3 Unlabeled-AM1A control screen*

A control screen of the ligands in the reaction buffer and with unlabeled AM1A was completed first in the fluorescence screening series to test for autofluorescence effects. The dialyzed AM1A stock was first thawed on ice and diluted into an intermediate solution containing 200 nM AM1A, 10 mM MOPS, pH = 6.5. This intermediate was denatured at  $90\text{ }^{\circ}\text{C}$  for 90 seconds and then renatured at room temperature over a half hour. A second solution was made to test for fluorescence effects in the reaction matrix containing 10 mM MOPS, pH = 6.5. The diluted ligand stocks stored at  $-20\text{ }^{\circ}\text{C}$  were thawed at  $4\text{ }^{\circ}\text{C}$  and stored on ice. These ligands were then diluted to an intermediate concentration of 0.1 mM. Lastly, a reaction buffer was prepared containing 50 mM MOPS, 50 mM NaCl, and 10  $\mu\text{M}$  EDTA, pH = 6.5. To test for ligand autofluorescence or fluorescence interaction with AM1A, the ligands and reaction buffer were mixed with AM1A such that final concentration was 10  $\mu\text{M}$  for the ligands, 100 nM AM1A, 55 mM MOPS (pH=6.5), 50 mM NaCl, and 0.01 mM EDTA, pH=6.5. The reaction matrix test set contained the same conditions, except no AM1A was included in the mixture. Reagents were added together to a final volume of 30  $\mu\text{L}$  in a 96-well mixing plate and placed on a shaker at 800 rpm for 1 minute. The samples were then transferred (20  $\mu\text{L}$ ) into a Corning (New York) 384-well low volume, round bottom plate and covered with AlumnaSeal. The plate was then briefly centrifuged to remove air bubbles and placed in the plate reader to incubate for 10 minutes at  $25\text{ }^{\circ}\text{C}$ , with the seal removed halfway through incubation.

Assayed wells were arranged on the plate such that each RNA model corresponded to single row and every other column and row was skipped to mitigate differences in fluorescence due to well location as seen in previous experiments. The TAMRA-labeled AM1A was excited at 543 nm and emission read at 590 nm in a Spectramax 5 (California) plate reader.

#### *8.4 5'-TAMRA-AM1A replicate screen*

Reproducibility of a binding response was analyzed by first completing a replicate study using 5'-TAMRA labeled AM1A. The AM1A stock was first thawed on ice and diluted into an intermediate solution containing 200 nM TAMRA-AM1A, 10 mM MOPS, pH = 6.5. This intermediate was aliquoted into two tubes, with one denatured at 90 °C for 90 seconds before renaturing at room temperature over a half hour while the other was stored on ice. The ligands were mixed with RNAs and reaction buffer (same concentrations as control assay) such that final concentration was 10 μM for the ligands, 100 nM TAMRA-AM1A, 55 mM MOPS (pH=6.5), 50 mM NaCl, and 0.01 mM EDTA, pH=6.5. Control wells contained the above mixture, but without any ligand. Reagents were added together to a final volume of 30 μL in a 96-well mixing plate and placed on a shaker at 800 rpm for 1 minute. The samples were then transferred (20 μL) into a Corning (New York) 384-well low volume, round bottom plate and covered with AlumnaSeal. The plate was then briefly centrifuged to remove air bubbles and placed in the plate reader to incubate for 10 minutes at 25 °C, with the seal removed halfway through incubation.

Assayed wells were organized such that each RNA model corresponded to single row and every other column and row was skipped to mitigate differences in fluorescence

due to well location as per previous experiments. The TAMRA-labeled AM1A was excited at 543 nm and emission read at 590 nm in a Spectramax 5 (California) plate reader. This procedure was completed a second time with the second aliquot of the TAMRA-AM intermediate to obtain and compare experiment replicate data.

#### *8.5 5'-TAMRA-RNA specificity screen and dose response*

Binding specificity was compared using 5'-TAMRA labeled RNA models. All antiterminator model RNAs were previously dialyzed and stored in 10 mM MOPS, pH=6.5 dilution buffer. The RNA models were thawed on ice and individually denatured at 90 °C for 90 seconds before renaturing at room temperature over a half hour. The denaturing and subsequent mixing and reading of each antiterminator model RNA was completed individually for optimal accuracy and to reduce risks of photobleaching. The ligands were mixed with RNAs and buffer such that final concentration was 10 µM for the ligands, 100 nM for the TAMRA-labeled RNA, 55 mM MOPS (pH=6.5), 50 mM NaCl, and 0.01 mM EDTA, pH=6.5. Control wells contained the above mixtures, but without any ligand. Reagents were added together to a final volume of 30 µL in a 96-well mixing plate and placed on a shaker at 800 rpm for 1 minute. The samples were then transferred (20 µL) into a Corning (New York) 384-well low volume, round bottom plate and covered with AlumnaSeal. The plate was then briefly centrifuged to remove air bubbles and placed in the plate reader to incubate for 10 minutes at 25 °C, with the seal removed halfway through incubation.

Assayed wells were organized such that each RNA model corresponded to single row and every other column and row was skipped to mitigate differences in fluorescence

due to well location as per previous experiments. The TAMRA-labeled RNA was excited at 543 nm and emission read at 590 nm in a Spectramax 5 (California) plate reader.

Fluorescence dose response assays followed the same protocol as the specificity screen, with multiple concentrations of a single compound tested with the antiterminator models (up to eight concentrations per row) instead of replicates of 10  $\mu$ M. The concentrations tested for amodiaquine were: 500  $\mu$ M, 100  $\mu$ M, 50  $\mu$ M, 25  $\mu$ M, 12.5  $\mu$ M, 2  $\mu$ M, and 100 nm. The concentrations tested for hydroxychloroquine were: 1 mM, 500  $\mu$ M, 100  $\mu$ M, 50  $\mu$ M, 25  $\mu$ M, 12.5  $\mu$ M, 2  $\mu$ M, and 100 nm. The compound concentrations were prepared in a dilution series.

#### *8.6 Ligand preparation for computational studies*

Compound libraries were collected and imported to the user documents folder at the Ohio Supercomputer Center (OSC). At the OSC, these files were exported to Maestro (Schrödinger) and minimized in their 3D structure using LigPrep. All ligands were prepared using the OPLS\_2005 forcefield in the pH range  $7.0 \pm 0.5$  to capture all protonation states near experimental conditions. Default settings in LigPrep were otherwise retained. This procedure was slightly modified for the ZINC library, with compounds only able to form produce up to 4 stereoisomers per compound to mitigate the computational cost of compounds with more than 2 chiral centers. LigPreps were completed using the Owens cluster at the OSC.

#### *8.7 Molecular descriptors*

Using the Molecular Descriptors application program in Schrödinger, the molecular descriptors of the compound library ligands were collected. The -out files from LigPrep (.mae) were exported to Maestro (Schrödinger) at the OSC. The QikProp

properties selected to identify were chirality count, Ruleoffive (Lipinski violations), solubility (either LogP or QPLogS), and Mol\_MW (molecular weight). No semiempirical calculations or topological properties were selected. Molecular descriptor jobs were completed using the Owens cluster at the OSC, and the -out files were exported to a personal MacBook to analyze using Maestro (academic license). Ligands with desirable descriptors were collected and imported to the OSC for use in molecular docking.

### *8.8 Molecular Docking*

Ligand dockings were completed using the Glide module in Maestro (Schrödinger). Ligands selected from the molecular descriptors protocol were docked to a grid of the NMR-solution structure of AM1A (Gerdeman et al., 2002). This grid does not contain the UUCG tetraloop at the top of Helix A2, which is not part of the conserved element and was therefore not a target for drug discovery efforts. The high throughput virtual screening (HTVS) mode was used in Schrödinger, with up to 5 ligand poses possible per compound. Per residue-interaction scores were selected to be written for each docking pose for analysis. No post-docking minimization was completed. Ligand dockings were completed using the Owens cluster at the OSC. Maestro out files (.maegz) were exported to a personal Macbook for analysis.

### *8.9 Interaction Fingerprinting*

Interaction Fingerprints in the Discovery Informatics and QSAR module of Maestro (Schrödinger) were used for analysis of ligand poses from computational docking experiments. Defaults were maintained, and the interaction fingerprint was written to the project table. This procedure was completed on a personal computer. The project table from this procedure was exported to a personal computer for analysis using

Microsoft Excel, as it contained molecular descriptor, docking, and interaction fingerprint information.

#### *8.10 Computational development of AMC model and Grid*

A model of AM<sub>Control</sub>, the bulge free model of the T-box antiterminator, was developed in Maestro, using the Build Biopolymer from Sequence function in the Build tool in Maestro. The model was built as an A-form RNA, as its structure follows standard Watson-Crick base pairs for all base-pairs except for the G-U wobble at the sixth position, and the secondary strand was automatically assigned by the software. The sequence input for the A strand was GAGGGGCGC, with the UUCG tetraloop excluded to match the AM1A structure. The structure was exported to PyMol at the OSC as a PDB file, and the complementary to G6 (C4 in the B strand of the RNA) was converted to a U using the Mutagenesis tool. This corrected AM<sub>Control</sub> structure was exported as a PDB file and loaded in Maestro for preparation.

The AM<sub>Control</sub> structure was optimized for grid preparation using the Protein Preparation Wizard in Maestro. No heteroatom states were generated in the preprocess, otherwise defaults were retained in the preprocess. Following preprocessing hydrogens and heavy atoms were refined. A Grid of this processed model was then developed using the Receptor Grid Generation program in the Glide module of Maestro. The receptor definition was deselected, aromatic hydrogens were selected to count as hydrogen bond donors, halogens were selected to count as halogen bond acceptors, and the OPLS\_2005 force field was selected in the Receptor tab. The grid was developed centroid to the RNA residues A2-G8 in the A chain and C2-U8 in the B chain, excluding both terminal base pairs, with the dock size  $\leq 20$  Å in the Site tab. Additionally the ligand diameter

midpoints were selected to lie within 18 Å of the center of the grid on all axes. The grid was then generated with all other settings default. Molecular docking to this AM<sub>Control</sub> model operated identically to the AM1A molecular dockings, with the AM<sub>Control</sub> grid selected instead.

## 9. Calculations and Data Analysis

### 9.1 Collection of Region Interactions

Following completion of Interaction Fingerprinting of a Molecular Docking procedure, the project table of the docking file was exported as a .csv file. This was imported to Microsoft Excel. For AM1A, compound interactions were totaled using the SUM function to collect all interactions for the helix A1 (residues A1-A5, B21-B25 in 1N53), the bulge (residues A6-A12), and helix A2 (residues A13-A16, B17-B20). These sums were added as additional columns to the project table for analysis of region selectivity for ligand poses.

### 9.2 Fluorescence Normalization

In the fluorescence reproducibility assay and specificity assay the measure of ligand strength, demonstrated through fluorescence modulation of 5'-TAMRA-labeled antiterminator models, was analyzed as a normalized value to allow swift comparison between compounds and models. Following import of the Softmax file of fluorescence data from an experiment to Excel, normalized fluorescence was calculated by the following formula:

$$F_N = \frac{\bar{F}_L}{\bar{F}_C}$$

Where  $F_N$  is the normalized fluorescence value,  $\bar{F}_L$  is the mean of all treated wells containing the ligand and the model RNA, and  $\bar{F}_C$  is the mean of the control wells, containing the reaction buffer and antiterminator model only. Normalized fluorescence values then used for plotting results in Graphpad Prism.

### 9.3 Error Propagation

In comparison of control well fluorescence as well as comparison of Glide Emodel values from computational results, the standard deviation of the mean for these values was suitable to demonstrate error. In the reproducibility assay and specificity assay however, standard deviation was not an appropriate indication of error for normalized fluorescence values, as two mean values were used in their calculation. Instead, the error of the means for the ligand well fluorescence values and control well fluorescence values were calculated by the following equation:

$$\delta F_n = F_N * \sqrt{\left(\frac{\delta F_L}{\bar{F}_L}\right)^2 + \left(\frac{\delta F_C}{\bar{F}_C}\right)^2}$$

Where  $\delta F_L$  is the standard deviation of ligand-containing mixture fluorescence values,  $\bar{F}_L$  is the mean of the ligand-containing mixture fluorescence values,  $\delta F_C$  is the standard deviation of control antiterminator model fluorescence values, and  $\bar{F}_C$  is the mean of the control fluorescence values. These error propagation values were calculated in Excel and were used in plotting the results of specificity and reproducibility screenings, displayed as error bars on the x-axis and y-axis using GraphPad Prism.

## 10. Bibliography

- About-ela, F. (2010) Strategies for the design of RNA-binding small molecules. *Future Med. Chem.* 2(1); 93-119.
- Acquaah-Harrison, G., Zhou, S., Hines, J., Bergmeier, S. (2010) Library of 1,4-disubstituted 1,2,3-triazole analogs of oxazolidinone RNA-binding agents. *J. Comb. Chem.* 12: 491-496.
- Aghdam, E., Barzegar, A., Hejazi, M. (2014) Evolutionary origin and conserved structural building blocks of riboswitches and ribosomal RNAs: riboswitches as probable target sites for aminoglycosides interaction. *Adv Pharm Bull.* 4(3):225-235.
- Al-Bari, M.A.A (2015) Chloroquine analogues in drug discovery: new directions of uses, mechanisms of actions and toxic manifestations from malaria to multifarious diseases. *J. Antimicrob. Chemother.* 70; 1608-1621.
- Antibiotic/Antimicrobial Resistance. *Centers for Disease Control and Prevention.* 2018
- Anupam, L., Denapoli, A., Muchenditsi, A., Hines, J.V. (2008) Identification of neomycin B-binding site in T box antiterminator model RNA R. *Bioorg. Med. Chem.*, 2008, 16, 4466-4470.
- Armstrong, A., Aldhumani, A.H., Schopis, J.L., et al. (2020) RNA drug discovery: Conformational restriction enhances specific modulation of the T-box riboswitch function. *Bioorg. Med. Chem.* 28; 115696.
- Battaglia, R., Grigg, J.C., Ke, A. (2019) Structural basis for tRNA decoding and aminoacylation sensing by T-box riboregulators. *Nat. Struct. Mol. Biol.* 26; 1106-1113.

- Brooks, W.H., Guida, W.C., Daniel, K.G. (2011) The Significance of Chirality in Drug Design and Development. *Curr. Top. Med. Chem.* 11(7); 760-770.
- Chetnani, B., and Mondragon, A. (2017) Molecular envelope and atomic model of an anti-terminated glyQS T-box regulator in complex with tRNA<sup>Gly</sup>. *Nucleic Acids Res.* 45: 8079–8090.
- Connelly, C.M., Moon, M.H., Schneekloth Jr., J.S. (2016) The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem. Biol.* 23(9); 1077-1090.
- Costales, M.G., Childs-Disney, J.L., Haniff, H.S., Disney, M.D. (2020) How We Think about Targeting RNA with Small Molecules. *J. Med. Chem.* 63, 8880-8900.
- Deigan, K., Ferré-D'Amaré. (2011) Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs. *Acc Chem Res.* 44(12):1329-1338.
- Disney, M.D., Winkelsas, A.M., Velagapudi, S.P., Southern, M., Fallahi, M. Childs-Disney, J.L. (2016) Inforna 2.0: a platform for the sequence-based design of small molecules targeting structured RNAs. *ACS Chem. Biol.* 11, 1720–1728.
- Fang, X., Michnicka, M., Zhang, Y., Wang, Y., and Nikonowicz, E. (2017) Capture and release of tRNA by the T-loop receptor in the function of the T-box riboswitch. *Biochemistry* 56: 3549–3558.
- Fauzi, H., Jack, K., Hines, J. (2005) In vitro selection to identify determinants in tRNA for bacillus subtilis tyrS T box antiterminator mRNA binding. *Nucleic Acids Res.* 8: 2595-2602.
- Friesner, R., Banks, J., Murphy, R., Halgren, T., Klicic, J. Mainz, D., Repasky, M., Knoll, E., Shelley, M., Perry, J., Shaw, D., Francis, P., Shenkin, P., (2004) Glide:

- a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47; 1739-1749.
- Frohlich, K., Weintraub, S., Bell, J., Todd, G., Vare, V., Schneider, R., Kloos, Z., Tabe, E., Cantara, W., Stark, C., Onwuanibe, U., Duffy, B., Basanta-Sanchez, M., Kitchen, D., McDonough, K., Agris, P. (2019) Discovery of small-molecule antibiotics against a unique tRNA-mediated regulation of transcription in gram-positive bacteria. *ChemMedChem.* 14; 758-769.
- Gerdeman, M., Henkin, T., and Hines, J. (2003) Solution structure of the bacillus subtilis T-box antiterminator RNA: seven nucleotide bulge characterized by stacking and flexibility. *J. Mol. Biol.* 326:189–201.
- Gerdeman, M., Henkin, T., Hines, J. (2002) In vitro structure-function studies of the bacillus subtilis tyrS mRNA antiterminator: evidence for factor-independent tRNA acceptor stem binding specificity. *Nucleic Acids Res.* 30:1065-1072.
- Green, N. J., Grundy, F. J., and Henkin, T. (2010) The T box mechanism: tRNA as a regulatory molecule. *FEBS Lett.* 584(2):318–324.
- Grundy, F., Rollins, S., and Henkin, T. (1994) Interaction between the acceptor end of tRNA and the T box stimulates antitermination in the bacillus subtilis tyrS gene: a new role for the discriminator base. *J Bacteriol.* 176: 4518–4526.
- Gutiérrez-Preciado A., Henkin T., Grundy F., Yanofsky C., Merino E. (2009) Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. *Microbiol. Mol. Biol. Rev.* 73(1):36-61.
- Haniff, H. S., Knerr, L., Liu, X., Crynen, G., Boström, J., Abegg, D., Adibekian, A., Lekah, E., Wang, K. W., Cameron, M. D., Yildirim, I., Lemurell, M., and Disney,

- M. D. (2020) Design of a small molecule that stimulates vascular endothelial growth factor A enabled by screening RNA fold–small molecule interactions. *Nat. Chem.* 12; 952–961.
- Henkin, T., Glass, B., Grundy, F. (1992) Analysis of the bacillus subtilis tyrS gene: conservation of a regulatory sequence in multiple tRNA synthase genes. *J. Bacteriol.* 174: 1299–1306.
- Hermann, T. (2016) Small molecules targeting viral RNA. *WIREs RNA.* 7:726-743.
- Hilimire, T.A, Chamberlain, J.M., Anokhina, V. (2017) HIV-1 Frameshift RNA-Targeted Triazoles Inhibit Propagation of Replication-Competent and Multi-Drug-Resistant HIV in Human Cells. *ACS Chem. Biol.* 12, 1674-1682.
- Hines, J., Means, J., Muchenditsi, A., Ageyman, A. (2010) Structural probing of the T box antiterminator-tRNA complex. *Biophysical J.* 98: S263a.
- Huggins, D.J., Sherman, W., Tidor, B. (2012) Rational Approaches to Improving Selectivity in Drug Design. *J. Med. Chem.* 55; 1424-1444.
- Irwin, J.J., Shoichet, B.K. (2005) ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model,* 45(1), 177-182.
- Jack, K., Means, J., Hines, J. (2008) Characterizing riboswitch function: identification of Mg<sup>2+</sup> binding site in T box antiterminator RNA. *Biochem. Biophys. Res. Commun.* 370: 306-310
- Jentzsh, F., Hines, J. (2011) Interfacing medicinal chemistry with structural bioinformatics: Implications for T box riboswitch RNA drug discovery. *BMC Bioinformatics.* 1(2);S5-S10.

- Juru, A.U., Hargrove, A.E (2020) Frameworks for targeting RNA with small molecules. *J. Biol. Chem.* 296, 100191.
- Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., Jorgensen, W. J. J. (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *Phys. Chem. B* 105, 6474-6487.
- Kreuzer, K.D., Henkin, T.M. The T box riboswitch: tRNA as an effector to modulate gene regulation. (2018) *Microbiol. Spectr.* 6 (4).
- Li, S., Su, Z., Lehmann, J., et al (2019) Structural basis of amino acid surveillance by higher-order tRNA-mRNA interactions. *Nat. Struct. Mol. Biol.* 26; 1094-1105.
- Lipinski, C (2004) Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.*
- Liu, J., Zeng, C., Hogan, V., Zhou, S., Monwar, Md.M., Hines, J. (2016) Identification of spermidine binding site in T box riboswitch antiterminator RNA. *Chem. Biol. Drug Design.* 87: 182-189.
- Liu, J., Zeng, C., Zhou, S., Means, J., Hines, J. (2015) Fluorescence assays for monitoring RNA-ligand interactions and riboswitch-targeted drug discovery screening. *Methods Enzymol.* 550:363-383.
- Means, J., Hines, J. (2005) Fluorescence resonance energy transfer studies of aminoglycoside binding to a T box antiterminator RNA. *Bioorg. Med. Chem. Letters.* 15: 2169-2172.

- Means, J., Katz, S., Nayek, A., Anupam, R., Hines, J., Bergmeier, S. (2006) Structure activity studies of oxazolidinone analogs as RNA-binding agents. *Bioorg. Med. Chem. Letters*. 16: 3600-3604.
- Means, J., Simson, C., Zhou, S., Rachford, A., Rack, J., Hines, J. (2009) Fluorescence probing of T box antiterminator RNA: insights into riboswitch discernment of the tRNA discriminator base. *Biochem. Biophys. Res. Commun.*, 389: 616-621.
- Morgan, B. S., Sanaba, B. G., Donlic, A., Karloff, D. B., Forte, J. E., Zhang, Y., and Hargrove, A. E. (2019) R-BIND: an interactive database for exploring and developing RNA-targeted chemical probes. *ACS Chem. Biol.* 14; 2691–2700.
- Morgan, B., Forte, J., Culver, N., Zhang, Y., Hargrove, A. (2017) Discovery of Key Physicochemical, Structural, and Spatial Properties of RNA-Targeted Bioactive Ligands. *Angew. Chem. Int. Ed.* 56; 13498-3502.
- Morgan, B.S., Forte, J.E., Culver, R.N., Zhang, Y., Hargrove., A.E. (2017) Discovery of Key Physicochemical, Structural, and Spatial Properties of RNA-Targeted Bioactive Ligands. *Angew. Chem. Int. Ed.* 56; 13498-13502.
- Padroni, G., Patwardhan, N., Schapira, M., Hargrove, A. (2020) Systematic analysis of the interactions driving small molecule–RNA recognition. *RSC Med Chem.* 11; 802-813.
- Putzer H., Condon C., Brechmeier-Baey D., Brito R., Grunberg-Manago D. (2002) Transfer RNA-mediated antitermination in vitro. *Nucleic Acids Res.* 30:3026–3033.

- Rizvi, N., Santa Maria Jr., J.P., Nahvi, A., et al. (2020) Targeting RNA with Small Molecules: Identification of Selective, RNA-Binding Small Molecules Occupying Drug-Like Chemical Space. *SLAS Discov.* 25(4); 384-396.
- Ryde, U., Söderhjelm, P. (2016) Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chem. Rev.* 116, 5520-5566.
- Shao, Y., Zhang, Q.C. (2020) Targeting RNA structures in diseases with small molecules. *Essays Biochem.* 64; 955-966.
- Shaw, J.S., Barbachyn, M.R. (2011) The oxazolidinones: past, present and future. *Ann. NY Acad. Sci.* 1241, 48-70.
- Stelzer, A. C., Frank, A. T., Kratz, J. D., Swanson, M. D., Gonzalez- Hernandez, M. J., Lee, J., Andricioaei, I., Markovitz, D. M., and Al- Hashimi, H. M. (2011) .Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* 7; 553–559.
- Sudarsan, N., Cohen-Chalamish, S., Nakamura, S., Emilsson, G.M., Breaker, R.R. (2005) Thiamine Pyrophosphate Riboswitches Are Targets for the Antimicrobial Compound Pyrithiamine. *Chem. Biol.* 12, 1325-1335.
- Suddala, K., Cabello-Villegas, J., Michnicka, M., Marshall, C., Nikonowicz, E., et al. (2018) Hierarchical mechanism of amino acid sensing by the T-box riboswitch. *Nat. Commun.* 9: 1896.
- Suddala, K., Zhang, J. (2019) An evolving tale of two interacting RNAs – themes and variations of the T-box riboswitch mechanism. *IUBMB.* 71(8): 1167-1180.
- Tran, T., Disney., M. (2012) Identifying the preferred RNA motifs and chemotypes that interact by probing millions of combinations. *Nat Commun* 3; 1125.

- Vitreschak, A., Mironov, A., Lyubetsky, V., Gelfand, M. (2008) Comparative genomic analysis of T-box regulatory systems in bacteria. *RNA*. 14: 717–735.
- Wang, J., and Nikonowicz, E. (2011) Solution structure of the K-turn and specifier loop domains from the bacillus subtilis tyrS t-box leader RNA. *J. Mol. Biol.* 408: 99–117.
- Wang, J., Henkin, T., and Nikonowicz, E. (2010) NMR structure and dynamics of the specifier loop domain from the bacillus subtilis tyrS T box leader RNA. *Nucleic Acids Res.* 38: 3388–3398.
- Warner, K.D., Hajdin, C.E., Weeks, K.M. (2018) Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* 17; 547-558.
- Yu, A.M., Choi, Y.H., Tu, M.J. (2020) RNA Drugs and RNA Targets for Small Molecules: Principles, Progress, and Challenges. *Pharmacol. Rev.* 72, 862-898.
- Zapp, M. L., Young, D. W., Kumar, A., Singh, R., Boykin, D. W., Wilson, W. D., and Green, M. R. (1997) Modulation of the Rev-RRE interaction by aromatic heterocyclic compounds. *Bioorg. Med. Chem.* 5; 1149–1155.
- Zhang, J., and Ferre-D'Amare, A. (2013) Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature*. 500:363–366.
- Zhang, J., Chetnani, B., Cormack, E., Alonso, D., Liu, W., et al. (2018) Specific structural elements of the T-box riboswitch drive the two-step binding of the tRNA ligand. *Elife*. 7: e39518.
- Zhang, J., Ferré-D'Amaré, A. (2015) Structure and mechanism of T-box riboswitches. *WIREs RNA*. 6: 419-433.

Zhou, S., Acquah-Harrison, G., Jack, K., Bergmeier, S., Hines, J. (2012) Ligand-induced changes in T box antiterminator RNA stability. *Chem. Biol. Drug Design.* 79: 202-208.

Zhou, S., Means, J.A, Acquah-Harrison, G., Bergmeier, S.C., Hines, J.V. (2012) Characterization of a 1,4-disubstituted 1,2,3-triazole binding to T box antiterminator RNA. *Bioorg. Med. Chem.* 20; 1298-1302.