

博士論文

Development of fluorogenic peptide probes carrying internally-
incorporated cyanine dyes as FID indicators for dsRNAs

(RNA 二重鎖を標的とする FID インジケーター：シアニン色素を
内部に組み込んだ蛍光性ペプチドプローブの開発)

Department of Chemistry, Graduate School of Science, Tohoku University

Lee En Ting Tabitha

2022

Acknowledgements

This thesis would not have been possible if not for the following people / organisations:

Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT) for granting me my scholarship, without which I would not have been able to study in Japan.

The IGPAS programme for granting me a position in Tohoku University and for Kohama-san and Kikuchi-san from the DiRECT office for always having my back.

Professor Iki Nobuhiko for introducing me to the Nishizawa Lab and personally calling up Nishizawa-sensei to secure a place for me.

Prof. Nishizawa Seiichi, for allowing me to join the Analytical Chemistry Laboratory and providing me with my amazing research topics and guidance.

Associate Prof. Sato Yuusuke, for being a brilliant scientist, with seemingly unexhaustive patience, who willingly supervised my research for the past five years. I truly cannot thank you enough.

The Nishizawa Lab members, both old and new, who have come alongside me to encourage me and cheer me on along the way. In my five years of being in Japan, I have learnt that it is no insignificant thing to be able to go to lab every day and not dread it. Being part of the Nishizawa lab has been incredibly special and I have been truly blessed. Every single lab member has in their own way supported me. Special mention goes out to Mogami Kenta, Yoshino Yukina, Kuwahara Kazuki, Chioma Okeke, Iwahashi Yuu, Suzuki Michiyuki, He Luo and Sakamoto Tomomi. I love you all!

Grace Center Church Sendai, my church family in Japan. You are the reason I came back to Sendai and I regret nothing. Thank you for teaching me the true sense of what family-in-Christ means. Thank you for all the prayers and practical love and support given to me. May more people come to Christ because of you and how you love each other! Special mention to the Muhling family, for all their love, counsel, and delicious food, to my Life Group – Nicola Erin Siegfried, Joanna Shutts, Sri Naga Sesa Karri – who have travelled with me through my ups and downs, and to Andrew Mays, for all the free coffee.

Timothy Jim, Michael Zielewski, Chip, my Singaporean friends (Evan, Wei Jie, Chloe, Sam and Greg), Jennifer – my incredible (and sometimes incredibly annoying) friends. Thank you for food and a listening ear to all my rants.

Audrey Stephanie, my lovely (but unfortunately not forever) roommate. Meeting you and having the privilege to be your roommate has been one of the best things about Japan. Having to do the rest of my life without you is honestly a bit daunting. Thank you for being my pillar of support, my personal chef, my gaming buddy, my fellow sister-in-Christ. You better come to Singapore.

Sean Kwek, my soon-to-be forever roommate. Thank you for sticking with me despite the distance. Thank you for all the little practical ways you support me and for all the times you point me back to Christ. I can't wait to do life with you.

My family. To my mother, for insisting that we facetime once a week. That has kept our family together and has helped me through so many times. Thank you also for paying for my plane tickets when I visit Singapore. To the brothers, for not forgetting about your only sister and keeping up to date with your lives. You know how special our family is and I never take that for granted.

My church family and friends back in Singapore, who have continued to keep me in their prayers all this time and who enthusiastically welcome me back home whenever I visit.

And lastly, Father God, who truly has been a Father to me in my time in Japan; who has disciplined, who has encouraged, who has guided, who has sent His Son to die for my sins, from who all blessings flow, both spiritual and physical. May Your name be praised forever. May my life point to how great You are.

Abstract

This thesis consists of four chapters.

In **Chapter 1**, I expounded on the dsRNA-targeting probe scene that serves as the backdrop for both Chapters 2 and 3. New analytical methods for double-stranded RNA (dsRNA) need to be developed to overcome problems such as poor selectivity or perturbation / destruction of the native dsRNA structure. To that end, I designed two fluorogenic peptide probes – the small molecule – PNA oligomer conjugate (SPOC) probe and the light-up peptide indicator (LUPI). These two probes both have a thiazole orange (TO) fluorophore that is internally-incorporated into the probe. Both probes were designed to be used as fluorescent indicators in a fluorescent indicator displacement (FID) assay to screen for new small molecules that could potentially act as drugs that target dsRNA. The mechanism for the FID assay as well as characteristics of good FID indicators were also discussed.

Chapter 2. The first dsRNA target we chose was the bacterial ribosomal RNA acceptor site (bac rRNA A-site). The bac rRNA A-site is the proofreading centre of the bacterial ribosome, and thus has been an attractive drug target for many decades. It contains an internal loop flanked by dsRNA regions. In this chapter, we designed the small molecule – PNA oligomer conjugate (SPOC) probe to simultaneously bind to the internal loop and the dsRNA region by a simple conjugation strategy between ATMND-C₂-NH₂ (internal loop binder) and a tFIT probe. The SPOC probe was expected to form a triplex with the bacteria rRNA A-site through Hoogsteen base-pairing, with the ATMND moiety binding to the internal loop.

This conjugation strategy, although simple, worked impressively. Notably, the SPOC probe was able to bind to the dsRNA target even at a neutral pH, which is an unfavourable condition for Hoogsteen base-pairing. This was because the conjugated ATMND moiety was able to act as an anchor, facilitating binding even at pH 7.0. The probe was measured to have a dissociation constant (K_d) of 190 ± 72 nM and bound specifically to the bac rRNA A-site over the corresponding human rRNA A-sites. Lastly, we showed how the SPOC probe had potential as an FID indicator by conducting a mock FID assay, where the degree of displacement was in good correlation with the reported K_d values of the test compounds.

However, the SPOC probe design has its drawbacks. The lack of information about small molecules that can specifically bind to RNA structural features as well as the need for a homopurine stretch of RNA for Hoogsteen-base pairing limits the dsRNA targets that SPOC can be applied to.

Thus, in **Chapter 3**, we turned to a new probe design – LUPI. This time, instead of targeting a dsRNA, we targeted a ribonucleoprotein (RNP) complex (RNA + RNA-binding protein, RBP). As such, LUPI was designed to be a peptide indicator based on a well-known RNP complex – the HIV-1 Tat protein – HIV-1 TAR RNA model. The Tat protein binds to the TAR RNA to greatly enhance viral transcription. Without the Tat-TAR interaction, gene expression and replication are adversely affected. Hence, finding Tat-TAR

inhibitors is an attractive therapeutic strategy. Unlike SPOC, LUPI is made up of amino acids, with its internally-incorporated TO acting as a surrogate amino acid, something that has never been reported before. Its amino acid sequence is based on the arginine rich motif (ARM) of the Tat protein, the part that is responsible for binding to TAR RNA.

By comparing LUPI with three other probes, one of which was the original Tat peptide fluorescent indicator, we showed that LUPI has superior light-up response, binding affinity ($K_d = 1.0 \pm 0.6$ nM) and selectivity towards the TAR RNA over two other unrelated RNAs. Moreover, in a mock FID assay with five test compounds, only the test compound with a K_d comparable to LUPI's was able to displace it, highlighting LUPI's ability to sieve out super-strong binders.

Finally, in **Chapter 4**, I summarized my findings for the SPOC and LUPI probes as well as provide some thoughts about how these two probes can be tailored for other specific dsRNA targets, as well as other possible ways to develop the probe designs even further.

Table of Contents

Chapter 1 – Introduction and Background	13
1. Introduction to Double-stranded RNA (dsRNA) and dsRNA Analytical Methods	15
2. Basic Introduction to SPOC (Chapter 2) and LUPI (Chapter 3)	18
2.1 Thiazole Orange	18
2.2 Internal Incorporation of Cyanine Dye	20
3. Overview of Drug Screening Assays	22
4. Fluorescent Indicator Displacement Assay	23
4.1 Characteristics of a Good Indicator	24
5. References	26
Chapter 2 – Small Molecule – PNA Oligomer for the Bacterial rRNA A-site	29
1. Introduction to Bacterial rRNA A-site	30
1.1 The Bacterial rRNA A-site	30
1.2 Aminoglycosides	32
2. Previous Research in Laboratory	34
2.1 ATMND-C ₂ -NH ₂	34
2.2 Triplex-forming Forced Intercalation Probe	35
3. Probe Design and Research Objectives	37
3.1 Small molecule – PNA Oligomer Conjugate Probe Design	37
3.2 Varying the Linker Length of ATMND-C _n -NH ₂	37
3.3 Research Objectives	38
3.4 Probe Sequence	38
4. Probe Synthesis	40
4.1 Synthesis of ATMND-C _n -NH ₂	40
4.2 Synthesis of SPOC Probes	40
5. Fluorescence Light-Up Experiments	42
5.1 Preliminary Fluorescence Experiments at pH 5.5	42

5.2 Fluorescence Experiments at pH 7.0	44
5.3 Melting Point (T_m) experiments to determine Binding Mode	45
5.4 Comparing the Fluorescence Response between the SPOC Probes	48
5.5 Reasons for Fluorescence Response at pH 7.0	49
5.5.1 ATMND- C_n -NH ₂ acts as an Anchor	49
5.5.2 Flexible Bac rRNA Structure facilitates Binding	50
6. Binding Affinity Experiments	53
7. Selectivity Experiments	55
8. Mock FID Assay	57
8.1 Investigating the effects of a Lower Concentration	58
8.2 Measuring the Limit of Detection of Neomycin	60
9. Conclusion and New Generations of SPOC	61
10. Experimental Methods	63
11. Supporting Information	71
12. References	75
Chapter 3 – Light-Up Peptide Indicator for HIV-1 TAR RNA	79
1. Introduction to RNPs	81
2. Previous Peptide Indicators	82
3. Probe Design and Research Objectives	83
3.1 Light-Up Peptide Indicator (LUPI) Design	83
3.2 Research Objectives	83
3.3 Choosing which Amino Acid to Substitute with TO	83
3.4 LUPI Amino Acid Sequence	84
3.5 Other Probes for Comparison	84
3.6 Comparing two types of Thiazole Orange	86
4. Significance of the HIV-1 Tat protein – TAR RNA RNP Complex	88
4.1 Binding of Tat protein to TAR RNA	88
5. Probe Synthesis	90
6. Fluorescence Light-Up Experiments	92

6.1 Comparing the two TOs	92
6.2 Comparing the Fluorescence Intensity amongst All the Probes	94
6.3 Fluorescence Quantum Yield of LUPI	95
7. Circular Dichroism Experiments	96
8. Binding Affinity Experiments	97
8.1 Limit of Detection of LUPI and TAR RNA	99
9. Selectivity Experiments	101
9.1 Salt Effect Experiments	103
9.1.1 Salt Effect Experiments of LUPI with TAR RNA	103
9.1.2 Salt Effect Experiments of LUPI with TAR RNA and Bac rRNA	105
10. Mock FID Assay	107
10.1 Introduction to the Test Compounds	107
10.2 Optimising the Conditions for the FID Assay	109
10.2.1 Optimising Incubation Time	109
10.2.2 Optimising the Number of Equivalents for the Test Compounds	111
10.2.3 Optimising the Concentrations of LUPI and TAR RNA	112
10.3 Optimised Mock FID Assay	114
11. Conclusion	116
12. Experimental Procedures	117
13. Supporting Information	121
14. References	132
Chapter 4 – Conclusion and Future Outlook.....	137
1. General Conclusion	139
2. Future Outlook for SPOC	140
3. Future Outlook for LUPI	141
5. References	142

Chapter 1

Introduction and Background

1. Introduction to Double-stranded RNA (dsRNA) and dsRNA Analytical Methods

In September 1957, Francis Crick, of Watson-Crick base-pairing fame, gave a lecture at the University College London where he shared bold claims that would eventually go down in history to become what is known as the “Central Dogma”. The present common understanding of the central dogma is that information from DNA is transferred to RNA which is then transferred to proteins. (An easier explanation, although not entirely faithful to Crick’s ideology, would be that ‘DNA makes RNA which makes proteins’). (Cobb, 2017) RNA was relegated to merely being a go-between for DNA and proteins.

65 years on, much research has been done to prove that RNA has functions other than just to be translated into proteins. 98.5% of the human genome is non-protein coding (Mattick, 2004). Yet ~85% of the genome is still translated into RNA (Hangauer, Vaughn & McManus, 2013). This sparked much interest in elucidating this pervasive transcription and discovering the functions of these non-coding RNAs. Since then, many sub-types of non-coding RNA have been discovered, such as small interfering RNA (siRNA), small nucleolar RNA (snoRNA), and Piwi-interacting RNA (piRNA), to name a few.

Adding to the complexity of how RNA functions is the presence of multiple types of secondary and tertiary structures that RNA can fold into. Unlike DNA which mostly stays in its stable double-helical structure, RNA can form hairpin loops, internal bulges and internal loops in its secondary structure, with a plethora of other tertiary structures. The structural aspects of RNA are important because often RNA derives its *function* from its *structure*. In other words, a region of highly-structured RNA is very likely to be a region with some function.

In the case of bacterial and viral RNA, regions of therapeutic significance often are *double-stranded RNA (dsRNA)* regions. It is important to make the distinction between single-stranded RNA (ssRNA) and dsRNA because the strategies to analyse each one are vastly different.

In the case of ssRNA, the unhybridized single-strand of nucleobases is free to pair with another strand of nucleotides through Watson-Crick base pairing. And hence sequence-selective analysis of ssRNA is easily achieved through oligonucleotide-based probes (eg. molecular beacons, fluorescence in situ hybridization (FISH), etc.).

Unfortunately, the same strategy cannot be used to analyse dsRNA as dsRNA does not contain nucleotides that are available for Watson-Crick base pairing. This has resulted in a much smaller number of methods for dsRNA analysis, with many scientists choosing instead to focus on ssRNA regions. However, as already mentioned, many RNAs of interest contain dsRNA, and not ssRNA, structures. Hence, it is essential, if not inevitable, that tools for dsRNA analysis must be further developed.

Existing methods to analyse dsRNA, as summarised in Fig. 2, include small molecules that bind to internal bulges or loops in dsRNA. (Sato et al., 2019; Watkins, Jiang, Nahar, Maiti & Arya, 2015; Jiang et al., 2015) The drawback of this method is that it only works for RNA that contains internal bulges/loops and that small molecules often bind non-specifically to other parts of the RNA. Immunological methods such as

using dsRNA-binding proteins or antibodies also have been developed (Kobayashi, Tomita & Sakamoto, 2009; Ku et al., 2020; Aramburu, Navas-Castillo, Moreno & Cambra, 1991) but these methods tend to be for the universal detection of all dsRNAs, and thus cannot discriminate between dsRNA of different sequences. The creative use of oligonucleotide helper probes to unwind the helical structure of dsRNA before hybridising the resultant single-stranded RNA with a fluorescent probe leads to sequence-specific detection (Qu et al., 2020), but inevitably destroys the native structure of the target RNA in doing so.

As such, there is a need for better dsRNA analytical methods that are not plagued by problems of non-specificity, while still maintaining the structure of the dsRNA.

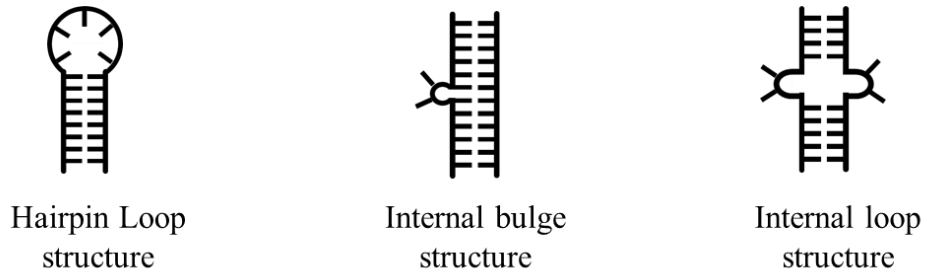


Figure 1. Example of secondary structures in dsRNA

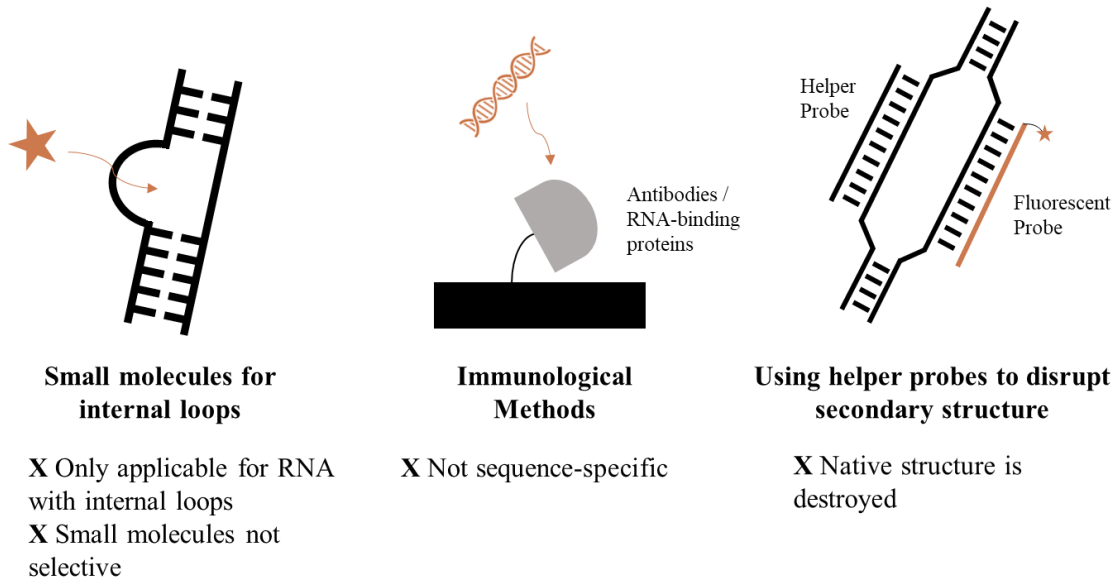


Figure 2. Summary of existing dsRNA analytical methods and the disadvantages associated with each type of method.

2. Basic Introduction to SPOC (Chapter 2) and LUPI (Chapter 3) Probes

In this thesis, I propose two novel dsRNA fluorescent probe designs to overcome the problems mentioned before. In Chapter 2, I designed the Small Molecule – PNA Oligomer Conjugate (SPOC) probe, while in Chapter 3, I developed the Light-Up Peptide Indicator (LUPI).

The SPOC probe employed a simple conjugation strategy between a small molecule (ATMND-C₂-NH₂) and a peptide nucleic acid (PNA) oligomer. Thiazole Orange (TO) was used as a surrogate base as well as a fluorophore.

LUPI is made up of amino acids based on a wild-type protein sequence, again using TO as its fluorophore and as a surrogate amino acid, which to the best of our knowledge, has never been done before.

Both SPOC and LUPI share two common features – the use of TO as well as the internal-incorporation of a cyanine dye.

2.1 Thiazole Orange (TO)

Thiazole orange is part of the family of cyanine dyes. Cyanine dyes typically have two nitrogen heterocyclic rings joined by a conjugated carbon system. Upon protonation, each of its resonance structures contains exactly one iminium group (Shindy, 2017).

TO consists of a benzothiazole group connected to a quinoline group via a monomethine bond. (Fig. 4A) When unbound, TO molecule can freely rotate around this monomethine bond, thus resulting in negligible fluorescence. However, upon intercalating into nucleic acids, TO's rotation is restricted and thus the loss of energy from its excited state is released as fluorescence. This gain of fluorescence is known as a light-up response, where the molecule lights up upon binding with the target. This light-up response is advantageous because it provides an excellent signal-to-noise ratio, allowing for more sensitive measurements to be made. This is also what sets TO apart from other nucleic acid intercalating dyes, such as pyrene (fluorescence quenched upon intercalation) (Østergaard & Hrdlicka, 2011).

Moreover, TO was shown by the Seitz group to be able to function as a surrogate nucleobase when paired against all four DNA nucleobases opposite it (Köhler, Jarikote & Seitz, 2004). We have also noticed that TO shows selectivity towards RNA over DNA, making it an ideal fluorophore to be used in dsRNA probes (Fig. 4C).

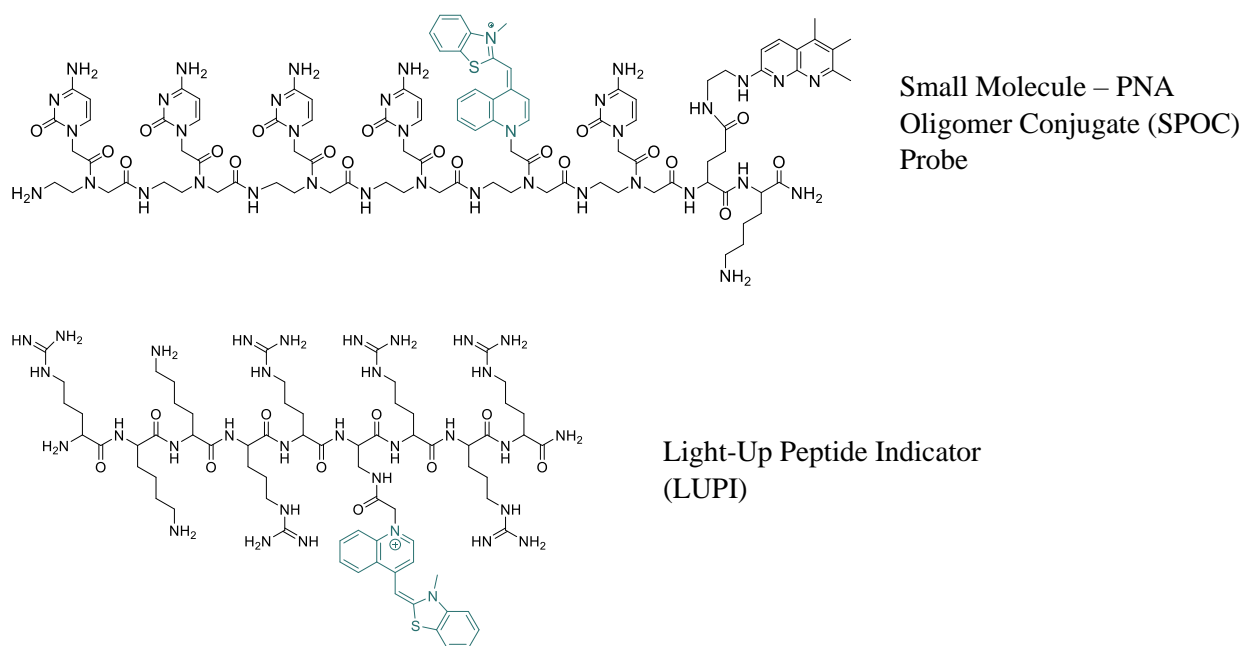


Figure 3. Chemical structures of SPOC and LUPI. The thiazole orange (TO) moiety in each probe is highlighted in a teal colour.

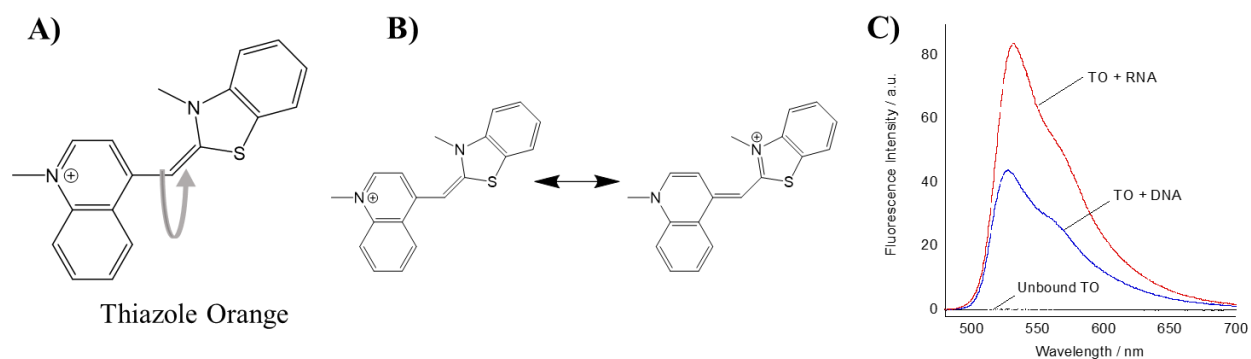
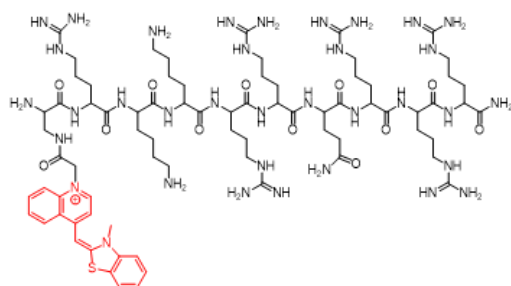


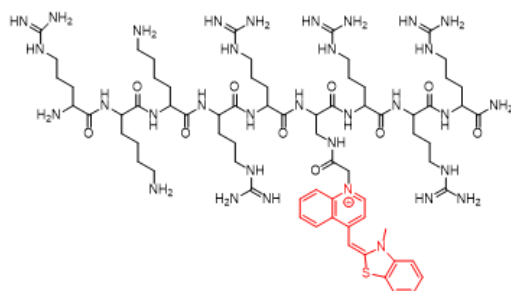
Figure 4. (A) Chemical structure of Thiazole Orange (TO). Rotation around the monomethine bond is shown with an arrow. (B) Resonance structures of TO. (C) Fluorescence spectrum of TO when unbound, and with RNA and DNA. [TO] = 1 μ M, [Nucleic Acid] = 1 mM, pH 7.0. RNA is *E. coli* total RNA and the DNA is ctDNA.

2.2 Internal Incorporation of Cyanine Dye

Both the SPOC probe and LUPI contain an internally-incorporated cyanine dye, which is in contrast with most conventional fluorescent probes which have their fluorophores connected at the terminus of the probe. The decision to connect the fluorophore in the middle of the chain of bases / amino acids was made because we predicted that this would lead to better intercalation of TO into the dsRNA, which would in turn result in a larger change in fluorescence. Furthermore, since TO was intended to be used as a surrogate, either as a surrogate nucleobase or surrogate amino acid, that meant that TO's position was not just limited to the ends of the probe but could be incorporated into the middle of the probe.



Terminally-linked
dye



Internally-
incorporated dye

Figure 5. Visual representation of the positional difference of the dye (red) when the dye is terminally-linked and when it is internally-incorporated.

3. Overview of Drug Screening Assays

The main application for both SPOC and LUPI was to be used in drug screening assays.

In recent years, dsRNA has become more popular as a drug target as traditional protein targets offer no new leads for effective antimicrobials. As such, more and more attention is being given to finding small molecule drug candidates that can bind well to a target RNA and thus trigger a therapeutic effect (Connelly, Moon & Schneekloth, 2016). To search for possible binders, large-scale screening of chemical libraries is an often-employed strategy.

Drug screening assays can be split into three main categories –affinity assays, *in silico* assays and mechanistic assays. (Martin, Grandi & Marcia, 2021) Affinity assays focus on the identification of direct ligand-target binding and includes techniques such as mass spectrometry and NMR spectroscopy. Surface plasmon resonance (SPR) has also been used in a high-throughput screen to identify binders (Lo et al., 2018). However, the binding of the test compound to the RNA might not necessarily result in a therapeutic effect, especially if the test compound binds non-specifically.

In silico assays are done using computer simulations and thus greatly increases the speed at which possible drugs can be found. *In silico* assays have been heavily used in searching for new drugs to target viral proteins. (Balmith, Faya & Soliman, 2016; Mehayar et al., 2021) They are gaining popularity for dsRNA targets whose secondary structures are well-resolved. One example is by Stelzer and his team, who calculated a dynamic ensemble of the HIV-1 TAR element RNA, an essential RNA for HIV replication, and virtually screened 51,000 small molecules against the model (Stelzer et al., 2011). However, as proteins have historically been the pharmaceutical targets, there lack of ligand-RNA binding data to form accurate predictions is a serious limitation.

Mechanistic, or phenotypic, assays focus on monitoring a mechanistic effect of the test compounds. A common mechanistic assay would thus be a cell-based assay, where the expression of a chemiluminescent or fluorescent protein (the mechanistic effect) would be used to determine if the test compound produced the desired effect or not. In cell-free contexts, mechanistic assays include fluorescence-based assays that can use a fluorescence resonance energy transfer (FRET) system (Bood et al., 2021), or an Alphascreen method (Mills, Shelat & Guy, 2007). The main disadvantage of mechanistic assays is that a test compound could possibly cause the mechanistic effect through other undesired mechanisms.

All these screening assays have their advantages and disadvantages. Hence, a robust drug identification process would include a primary screen, secondary screen(s) and compound optimisation. This research focuses on a primary screening technique known as the fluorescent indicator displacement assay.

4. Fluorescent Indicator Displacement Assay

The fluorescent indicator displacement (FID) assay involves the displacement of a fluorescent indicator to sieve out hit compounds.

In the case of dsRNA targets, a fluorescent indicator is first hybridised with the dsRNA and at the right excitation wavelength, the indicator will fluoresce. If a test compound (drug candidate) that binds very strongly to the dsRNA target is added, the test compound will displace the fluorescent indicator, resulting in an observable change in fluorescence intensity (Fig. 6).

The advantages of using an FID assay are that it is incredibly easy to perform and the use of fluorescence allows for a sensitive assay, as compared to other detection techniques such as anisotropy. There is no need to label the RNA and thus risk possible perturbations in the native structure of the RNA (Zhang, Umemoto & Nakatani, 2010). The FID assay is also readily adaptable to high-throughput applications by using a 96-well plate.

However, the major drawback of an FID assay is that the discriminatory ability of the assay relies solely on the indicator. A bad indicator would not be able to discriminate between strong-binding test compounds and weak-binding test compounds. For an FID assay to operate well, a good indicator is needed.

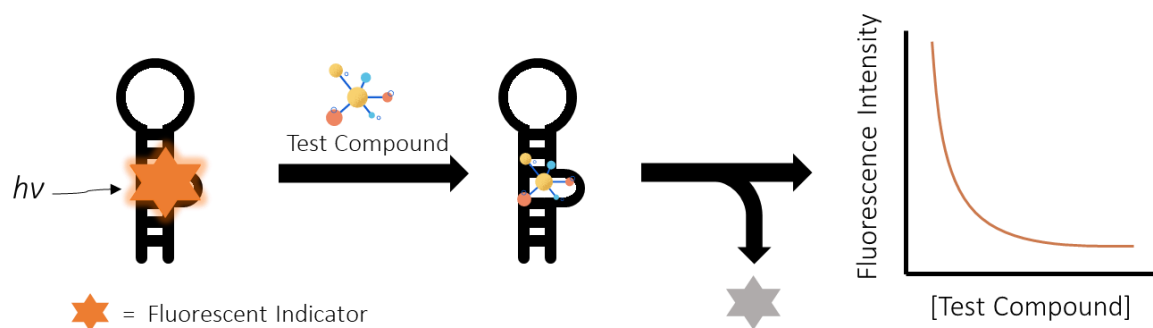


Figure 6. Schematic of a Fluorescent Indicator Displacement (FID) assay.

4.1 Characteristics of a Good Indicator

We have determined three qualities that a good FID indicator should have.

Firstly, a good indicator should have a **large fluorescence response** upon binding with the target RNA. The larger the change in fluorescence, the larger the range of displacement, the more sensitive the assay can be. For example, if the indicator has a very high background signal and the fluorescence intensity only increases twofold upon binding with the dsRNA, there is only a narrow range of onefold of the fluorescence intensity by which displacement can be observed. Thus, more subtle differences in the binding abilities of the various test compounds would not be able to be detected. The larger the change in fluorescence response, the more sensitive the FID assay.

Secondly, a good indicator should have **good binding affinity with the target RNA** so that only test compounds that have similar or better binding affinities will be able to displace it. An indicator that can be displaced by even weak-binding test compounds would be unfit for screening. Thus, it is important that the binding affinity of the indicator be tuned to match the desired binding affinity of hit compounds. For example, if objective is to screen for test compounds with super-strong binding, then it follows that the indicator itself must possess super-strong binding affinity. However, if the objective of the screen is to find test compounds with modest binding affinity, then the indicator's binding affinity should match that too. In most cases, it is assumed that good drug candidates would have strong binding affinities with the drug target, as such, we have decided that a good indicator should have good binding affinity with the target RNA.

Lastly, a good indicator should be **selective to the drug target**. For example, if the dsRNA contains an internal loop and if it has been reported that the internal loop is important for dsRNA's function, then the indicator should selectively bind to the internal loop and not other parts of the dsRNA, or to other dsRNAs. This is because if the indicator binds to some other part of the dsRNA and is eventually displaced by a test compound, it would result in a false positive as the decrease in fluorescence intensity is due to off-binding and not actual binding with the region of significance. As such, we strongly believe that a good indicator should be selective to the drug target, and more importantly, the region of significance (eg. the internal loop within the dsRNA).

1.5 Overall Research Objective

Considering all that has been written, the research objective of my thesis is as such:

To design new fluorogenic peptide probes for dsRNA analysis that contain an internally-incorporated cyanine dye for the use as indicators in FID assays.

To this end, I have designed the small molecule – PNA oligomer conjugate (SPOC) probe in Chapter 2 as well as the light-up peptide indicator (LUPI) in Chapter 3 and have evaluated their use as FID indicators (Fig. 7).

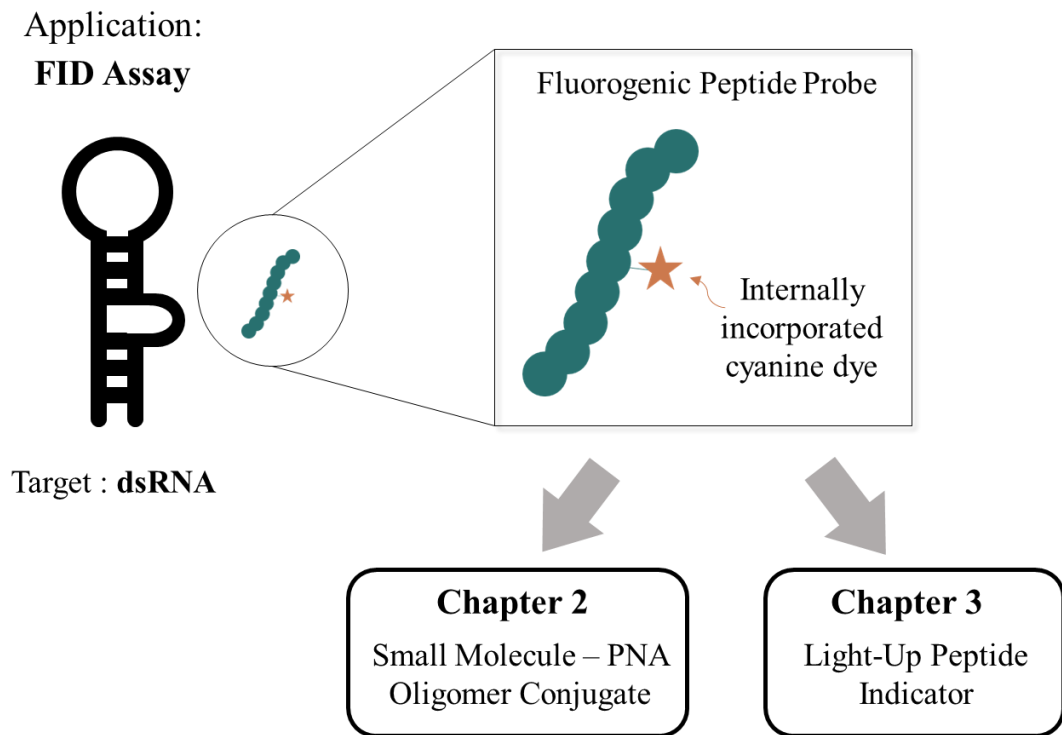


Figure 7. Summary of common features of probe designs and overall research objective.

5. References

- Aramburu, J., Navas-Castillo, J., Moreno, P., & Cambra, M. (1991). Detection of double-stranded RNA by ELISA and dot immunobinding assay using an antiserum to synthetic polynucleotides. *Journal Of Virological Methods*, 33(1-2), 1-11.
- Balmith, M., Faya, M., & Soliman, M. (2016). Ebola virus: A gap in drug design and discovery - experimental and computational perspective. *Chemical Biology & Drug Design*, 89(3), 297-308.
- Bood, M., del Nogal, A., Nilsson, J., Edfeldt, F., Dahlén, A., & Lemurell, M. et al. (2021). Interbase-FRET binding assay for pre-microRNAs. *Scientific Reports*, 11(1), 9396.
- Cobb M (2017) 60 years ago, Francis Crick changed the logic of biology. *PLoS Biology* 15(9): e2003243.
- Connelly, C., Moon, M., & Schneekloth, J. (2016). The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chemical Biology*, 23(9), 1077-1090.
- Hangauer, M., Vaughn, I., & McManus, M. (2013). Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *Plos Genetics*, 9(6), e1003569.
- Jiang, L., Watkins, D., Jin, Y., Gong, C., King, A., & Washington, A. et al. (2015). Rapid Synthesis, RNA Binding, and Antibacterial Screening of a Peptidic-Aminosugar (PA) Library. *ACS Chemical Biology*, 10(5), 1278-1289.
- Kobayashi, K., Tomita, R., & Sakamoto, M. (2009). Recombinant plant dsRNA-binding protein as an effective tool for the isolation of viral replicative form dsRNA and universal detection of RNA viruses. *Journal Of General Plant Pathology*, 75(2), 87-91.
- Köhler, O., Jarikote, D., & Seitz, O. (2004). Forced Intercalation Probes (FIT Probes): Thiazole Orange as a Fluorescent Base in Peptide Nucleic Acids for Homogeneous Single-Nucleotide-Polymorphism Detection. *Chembiochem*, 6(1), 69-77.
- Ku, J., Kim, S., Park, J., Kim, T., Kharbash, R., & Shin, E. et al. (2020). Reactive Polymer Targeting dsRNA as Universal Virus Detection Platform with Enhanced Sensitivity. *Biomacromolecules*, 21(6), 2440-2454.
- Lo, C., Li, O., Tang, W., Hu, C., Wang, G., & Ngo, J. et al. (2018). Identification of influenza polymerase inhibitors targeting C-terminal domain of PA through surface plasmon resonance screening. *Scientific Reports*, 8(1), 2280.
- Martin, W., Grandi, P., & Marcia, M. (2021). Screening strategies for identifying RNA- and ribonucleoprotein-targeted compounds. *Trends In Pharmacological Sciences*, 42(9), 758-771.
- Mattick, J. (2004). RNA regulation: a new genetics?. *Nature Reviews Genetics*, 5(4), 316-323.

- Mehyar, N., Mashhour, A., Islam, I., Gul, S., Adedeji, A., Askar, A., & Boudjelal, M. (2021). Using in silico modelling and FRET-based assays in the discovery of novel FDA-approved drugs as inhibitors of MERS-CoV helicase. *SAR And QSAR In Environmental Research*, 32(1), 51-70
- Mills, N., Shelat, A., & Guy, R. (2007). Assay Optimization and Screening of RNA-Protein Interactions by AlphaScreen. *SLAS Discovery*, 12(7), 946-955.
- Qu, G., Sun, X., Ying, N., Bu, S., Li, Z., & Hao, Z. et al. (2020). 16S rRNA-functionalized multi-HCR concatemers in a signal amplification nanostructure for visual detection of *Salmonella*. *Biotechnology And Applied Biochemistry*, 68(3), 560-567.
- Shindy, H. (2017). Fundamentals in the chemistry of cyanine dyes: A review. *Dyes And Pigments*, 145, 505-513.
- Stelzer, A., Frank, A., Kratz, J., Swanson, M., Gonzalez-Hernandez, M., & Lee, J. et al. (2011). Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nature Chemical Biology*, 7(8), 553-559.
- Watkins, D., Jiang, L., Nahar, S., Maiti, S., & Arya, D. (2015). A pH Sensitive High-Throughput Assay for miRNA Binding of a Peptide-Aminoglycoside (PA) Library. *PLOS ONE*, 10(12), e0144251.
- Østergaard, M., & Hrdlicka, P. (2011). Pyrene-functionalized oligonucleotides and locked nucleic acids (LNAs): Tools for fundamental research, diagnostics, and nanotechnology. *Chemical Society Reviews*, 40(12), 5771.
- Zhang, J., Umemoto, S., & Nakatani, K. (2010). Fluorescent Indicator Displacement Assay for Ligand-RNA Interactions. *Journal of The American Chemical Society*, 132(11), 3660-3661.

Chapter 2

Small Molecule – PNA Oligomer for the Bacterial rRNA A-site

En Ting Tabitha Lee, Yusuke Sato, Seiichi Nishizawa (2020). Small Molecule-PNA Oligomer Conjugates for rRNA A-site at neutral pH for FID Assays. *Chemical Communications*, 56, 14976-14979. (Highlighted at Front Cover)

西澤精一, LEE En Ting Tabitha, 芳野幸奈, 矢島さやか, 六川正文, 佐藤雄介 (2021). rRNA を標的とする蛍光プローブの分子設計：FID法における蛍光インジケータと生細胞核小体イメージング蛍光色素" *分析化学*, **70**, 703-714. (総合論文：「第81回分析化学討論会」特集号)

西澤精一, 佐藤貴哉, LEE En Ting Tabitha, 坂本直柔, 千葉年輝, 田邊貴昭, 芳野幸奈, 高橋勇樹, 佐藤雄介 (2022). シアニン色素擬塩基を有する三重鎖形成ペプチド核酸プローブによる RNA 二重鎖構造蛍光検出. *分析化学*, **71**, 133-144. (総合論文：分析試薬研究懇談会特集号)

1. Introduction to Bacterial rRNA A-site

1.1 The Bacterial rRNA A-site

The ribosome is the protein-making factory of a cell and is made up of ribosomal proteins and ribosomal RNA (rRNA). The ribosome is a universally conserved enzyme that exists in both prokaryotic and eukaryotic cells (Melnikov et al., 2012). As such, the prokaryotic bacterial ribosome has long been a target for antibiotics. This is especially true of the 16S rRNA A-site (acceptor / aminoacyl site) inside the bacteria's ribosome. Located in the 30S small subunit of the ribosome, the A-site serves as the proof-reading centre for the synthesis of polypeptides. Transfer RNAs (tRNAs) enter the ribosome through the A-site and if the anti-codon of the tRNA matches the messenger RNA (mRNA) that is being decoded, the tRNA is accepted and its amino acid is added to the polypeptide chain (Ramakrishnan, 2002). Upon binding with a cognate (ie. correct) tRNA, two adenosine residues in the A-site, A1492 and A1493 (following *E. coli*'s numbering), will flip out of the internal loop. However, this conformational change does not happen for non-cognate (ie. incorrect) tRNAs, thus forming the proof-reading mechanism that allows the ribosome to discriminate between cognate and non-cognate tRNAs (Ogle et al., 2001).

Structural features of the A-site rRNA include an internal loop (containing the three adenines that make up the A-site) as well as two flanking double-stranded RNA (dsRNA) regions (Fig. 1). Since the conformational change of the internal loop region governs the proofreading ability of the A-site, most drug strategies targeting the bacterial ribosome focus would focus on the A-site being the drug binding site (Barbieri, Kaul & Pilch, 2007).

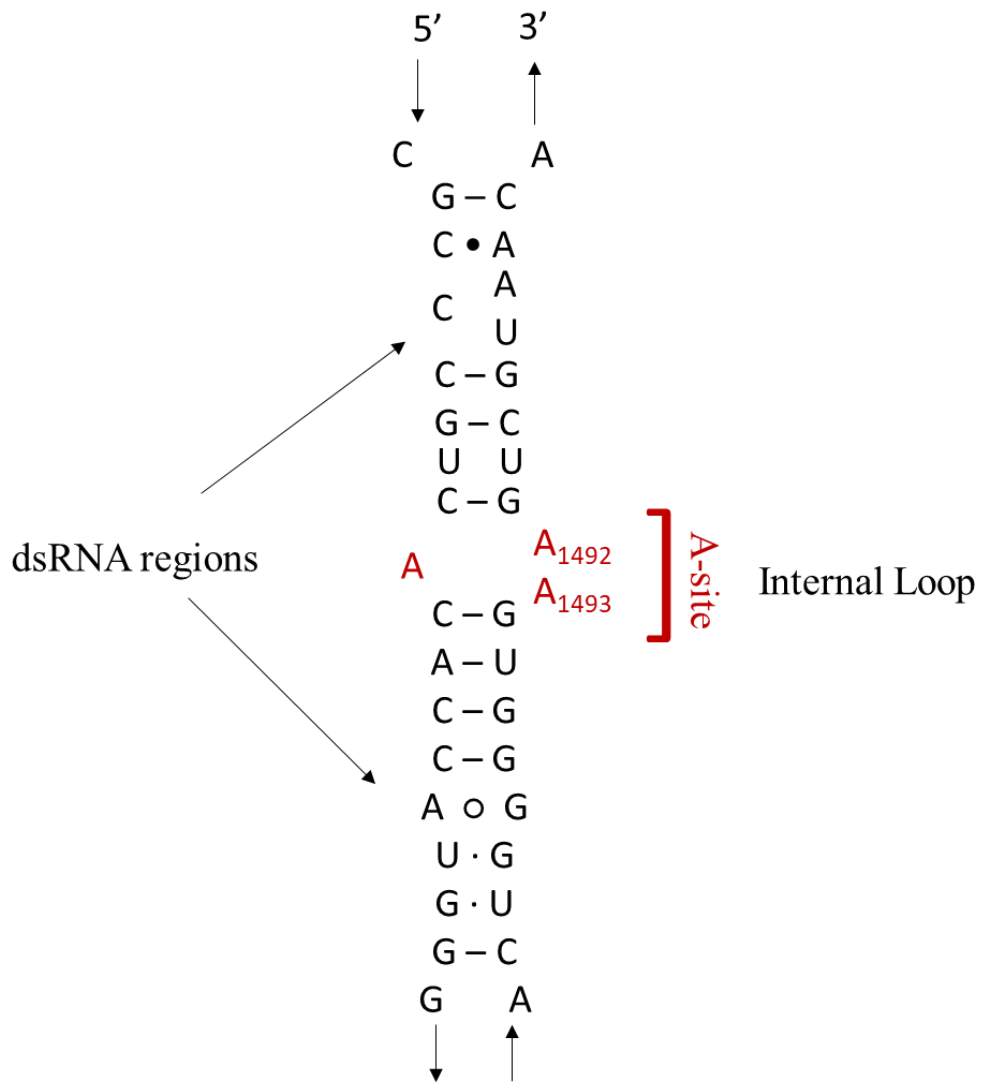


Figure 1. RNA sequence of the *E. coli* A-site rRNA. The structural features of the RNA have also been annotated.

1.2 Aminoglycosides

Aminoglycosides are a class of broad-spectrum antibiotics that are known to, amongst other things, bind to the bacterial A-site rRNA (Jana & Deb, 2006). The very first aminoglycoside was discovered in 1944 and was isolated from bacteria called *Streptomyces griseus*. This antibiotic was named streptomycin, and it became one of the most effective drugs against a broad spectrum of gram-negative bacterial infections, such as tuberculosis (Mingeot-Leclercq, Glupczynski & Tulkens, 1999). Since then, more aminoglycosides have been discovered and/or synthesised, including paromomycin, neomycin, kanamycin and tobramycin. These compounds share a similar structure, having several, usually three, cyclitol rings with amino or hydroxyl functional groups which are linked together via glycosidic bonds (eg. Fig. 2) (Forge & Schacht, 2000).

Although aminoglycosides have been used clinically since 1944, their mechanism of action was not precisely known. Scientists knew that aminoglycosides could bind to the rRNA A-site and stabilise the near-cognate codon recognition complex (Pape, Wintermeyer & Rodnina, 2000), but they were not sure how exactly this happened. It was only in the year 2000, when Carter and his team published a paper showing the crystal structure of the 30S ribosomal subunit bound to paromomycin that the mechanism was elucidated (Carter et al., 2000). His research showed that aminoglycosides, such as paromomycin, bind to the A-site and induce the same flip-out conformational change as in the case of a cognate tRNA. This tricks the ribosome into accepting non-cognate tRNAs, which increases the misincorporation of amino acids into the peptide chain (Fig. 2), resulting in dysfunctional proteins, which eventually leads to the death of the bacteria (Krause, Serio, Kane & Connolly, 2016; Lin, Zhou, Steitz, Polikanov & Gagnon, 2018).

Aminoglycosides, however, are not without their drawbacks. They are known to also bind non-specifically to other parts of the RNA, leading to toxic side effects (Hong et al., 2015). Furthermore, in the recent years, there has been a growing bacterial resistance to aminoglycosides (Chittapragada, Roberts & Ham, 2009; Macmaster, Zelinskaya, Savic, Rankin & Conn, 2010). One of the main mechanisms of resistance is the use of aminoglycoside-modifying enzymes that inactivate aminoglycosides so that they can no longer bind to the rRNA A-site (Garneau-Tsodikova & Labby, 2016). While aminoglycosides are still being used today, they are being used much more sparingly in a bid to slow down the increase of antimicrobial resistance. This has accelerated the hunt for new antibiotic candidates, especially since the bacterial A-site still proves to be a promising antibacterial target.

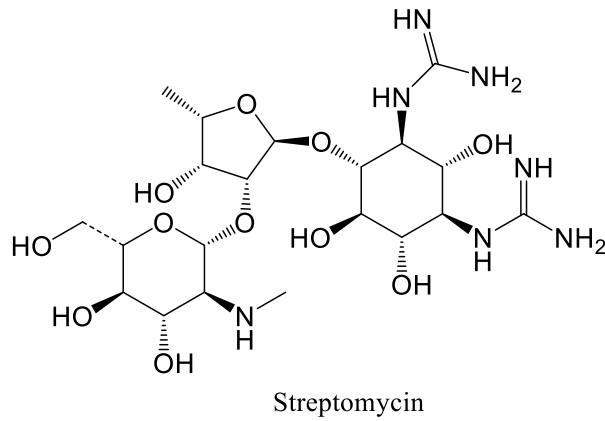


Figure 2. Chemical structure of streptomycin, the first aminoglycoside to be discovered.

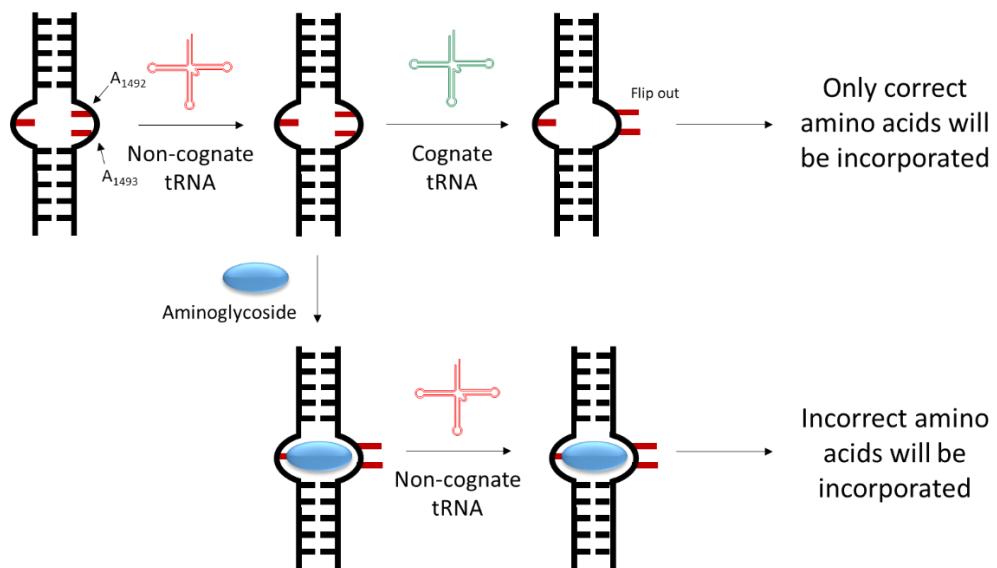


Fig. 3 Scheme of how aminoglycosides cause the ribosome to accept non-cognate tRNAs.

2. Previous Research in Laboratory

2.1 ATMND-C₂-NH₂

Our laboratory has previously discovered that a small molecule, ATMND-C₂-NH₂ (Fig. 4), exhibits the strongest binding to the A-site RNA out of all non-aminoglycosidic ligands ($K_d = 0.44 \mu\text{M}$, 5°C) (Sato et al., 2018). It does this by binding to the internal loop within the A-site, and its two-ring system allows for favourable π -stacking interactions. Moreover, upon binding to the A-site, ATMND-C₂-NH₂ does not drastically alter the structure of the A-site. However, ATMND-C₂-NH₂ itself suffers from the same lack of specificity as the aminoglycosides and is unable to distinguish between bacterial and human A-site rRNAs.

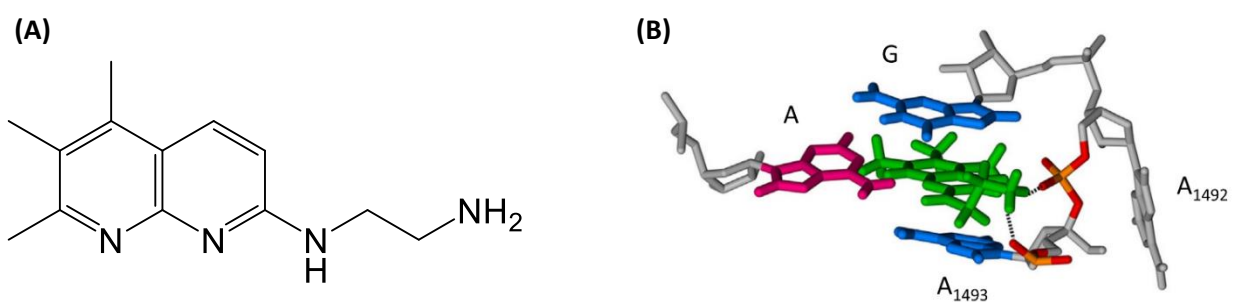


Fig. 4 (A) Chemical structure of ATMND-C₂-NH₂. (B) Side view of the possible binding mode to bacterial A-site rRNA. ATMND-C₂-NH₂ is pictured in green. Reproduced with permission from Sato, Y.; Rokugawa, M.; Ito, S.; Yajima, S.; Sugawara, H.; Teramae, N.; Nishizawa, S. Fluorescent Trimethylated Naphthyridine Derivative With An Aminoalkyl Side Chain As The Tightest Non-Aminoglycoside Ligand For The Bacterial A-Site RNA. *Chemistry - A European Journal*, 2018, 24 (52), 13862-13870. Copyright 2018, John Wiley and Sons.

2.2 Triplex-forming Forced Intercalation Probe

Our laboratory has also developed the triplex-forming forced intercalation (tFIT) probe (Sato, Sato & Nishizawa, 2016). This probe is made up of peptide nucleic acid (PNA) with thiazole orange as its fluorophore and a surrogate nucleobase (Köhler, Jarikote & Seitz, 2004). This fluorogenic PNA probe binds to dsRNA targets through Hoogsteen base-pairing to form a complex of three strands (ie. a triplex) (Fig. 5,6). Furthermore, it was shown that tFIT probes show a preference for RNA with a high binding affinity ($K_d = 23$ nM, as compared to 870 nM for DNA), making them very suitable as dsRNA-targeting probes. tFIT probes are designed to bind tightly to RNA strands not only through complementary Hoogsteen base-pairing, but also through suitable electrostatic factors. Using neutrally-charged PNA in tFIT probes instead of negatively-charged DNA / RNA allows for less repulsion with the RNA target and thus a higher binding affinity. TO itself is singly protonated under neutral conditions (Fig. 5), which would increase electrostatic attraction between the probe and the target.

However, for Hoogsteen base-pairing to occur, the cytosine involved in Hoogsteen base-pairing must be protonated. Since cytosine has a pK_a of ~ 4.4 (Pack, Wong & Lamm, 1998), this means that an acidic environment is required for Hoogsteen base-pairing to take place.

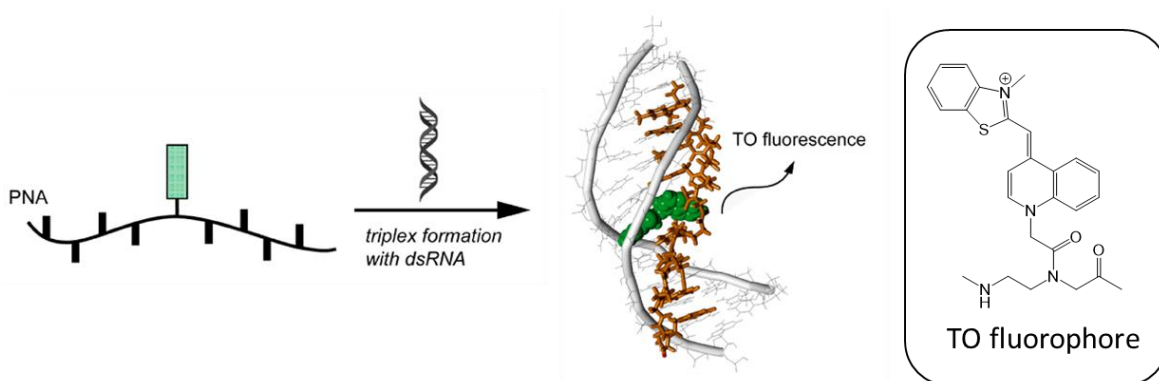


Figure 5. Schematic of a tFIT probe and how it hybridises with dsRNA to form a triplex. TO's chemical structure is included in the boxed region. Reprinted (adapted) with permission from *J. Am. Chem.*

Soc. **2016**, 138, 30, 9397–9400. Copyright 2016 American Chemical Society.

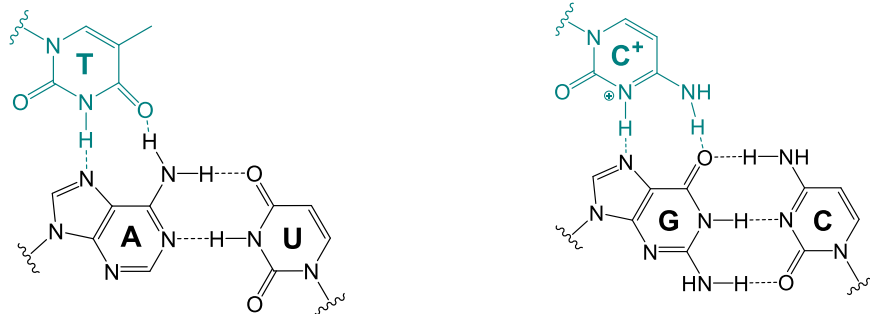


Figure 6. Examples of Hoogsteen-base-pairs. The nucleobases involved in Hoogsteen-base-pairing are shown in teal.

3. Probe Design and Research Objectives

3.1 Small molecule – PNA Oligomer Conjugate Probe Design

Since ATMND-C₂-NH₂ can target the internal loop region, and the tFIT probe can target the dsRNA region, we hypothesised that a simple conjugation between ATMND-C_n-NH₂ and a tFIT probe would result in a triplex-forming indicator that shows even tighter binding to the A-site RNA as well as a specificity toward the bacterial rRNA A-site through the sequence-selectivity conferred by the PNA component (Fig. 7). In doing so, the conjugated probe would be able to bind simultaneously to both the internal loop (with the ATMND-C₂-NH₂ moiety) and the complimentary RNA strand (with the tFIT moiety). We called this probe the small molecule – PNA oligomer conjugate (**SPOC**) probe.

3.2 Varying the Linker Length of ATMND-C_n-NH₂

To optimise the binding affinity of the SPOC probe, we decided to vary the linker length of the ATMND moiety when it was coupled to the tFIT moiety. As such, we synthesised three versions of the SPOC probe, where the ATMND-C_n-NH₂, where n = 2, 3 and 4. The resultant probes were named SPOC (n=2,3 or 4, according to the linker length).

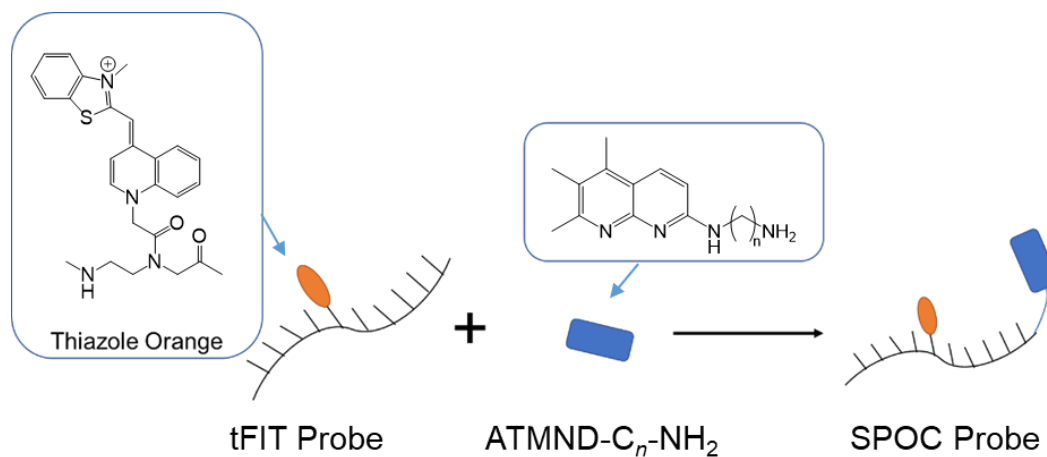


Figure 7. Simple conjugation strategy of a small molecule (ATMND-C_n-NH₂) with a tFIT probe to form a SPOC probe.

3.3 Research Objectives

The objective for SPOC was that the conjugation strategy would lead to a probe that:

1. Has a higher binding affinity toward the bacterial rRNA A-site compared to its individual components (the tFIT probe or the ATMND moiety)
2. Exhibits better selectivity toward the bacterial rRNA A-site over the human rRNA equivalents
3. Can be used as an FID indicator

As such, the experiments conducted were done with these objectives in mind.

3.4 Probe Sequence

The tFIT moiety sequence was designed to be complementary to a 6-base pair region of the bacterial rRNA A-site right below the internal loop. This was decided based on two reasons: firstly, that Hoogsteen base-pairing required a homopurine stretch of nucleobases for the probe to hybridise to, and secondly, that the internal loop had to be close to the targeted dsRNA sequence such that the tethered ATMND-C_n-NH₂ could still bind to the internal loop. In the chosen dsRNA sequence, there is a pyrimidine present (uracil), which cannot undergo Hoogsteen base-pairing. To overcome this problem, we chose to position TO opposite the uracil as TO can act as a universal surrogate base, thus mitigating the problem of a mismatched base-pair. A Lys residue was also added to the C' terminal of all the probes to increase solubility and stability of the probe-RNA complex (Ray & Nordén, 2000). A control probe that did not contain the ATMND moiety, and thus acted as normal tFIT probe, was also synthesised for comparison. For the control probe, to match the overall length of the probe with the SPOC probes, the Glu residue was substituted with a Gly residue.

Table 1. Table of all the probe sequences

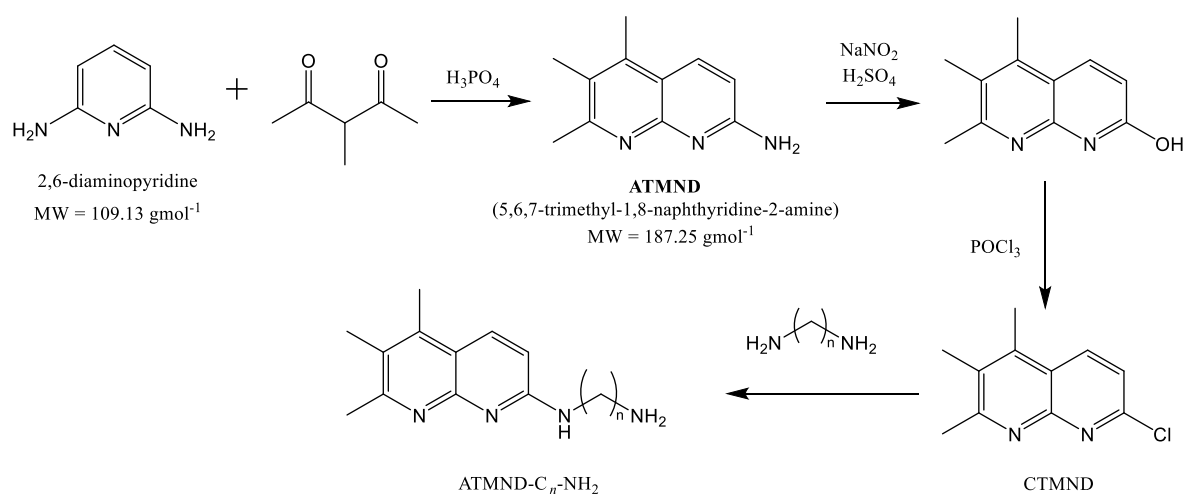
Probe	Probe Sequence (C' – N')
SPOC (n=2)	Lys Glu (ATMND-C ₂ -NH ₂) C (TO) C C C C
SPOC (n=3)	Lys Glu (ATMND-C ₃ -NH ₂) C (TO) C C C C
SPOC (n=4)	Lys Glu (ATMND-C ₄ -NH ₂) C (TO) C C C C
Control Probe	Lys Gly C (TO) C C C C

4. Probe Synthesis

4.1 Synthesis of ATMND- C_n -NH₂

ATMND- C_2 -NH₂ was synthesised according to previous literature (Wang, Sato, Kudo, Nishizawa & Teramae, 2012). The other ATMND versions were similarly synthesised as per the scheme below.

The general synthesis scheme is as follows:



Scheme 1. Overall synthesis scheme of ATMND- C_n -NH₂.

4.2 Synthesis of SPOC Probes

The SPOC probes were synthesised using solid-phase peptide synthesis using Fmoc chemistry.

The detailed protocol is written under the Experimental Procedures.

Thiazole orange was coupled to a PNA backbone monomer.

To couple ATMND- C_n -NH₂ to the PNA oligomer, we tried both Glu and Asp amino acids. However, the probe with Asp amino acid failed because aspartic acid had the tendency to form aspartimide (Neumann, Farnung, Baldauf & Bode, 2020). Hence, Glu was chosen as the linker for ATMND- C_n -NH₂.

The synthesised probes were purified using HPLC and confirmed using MALDI-TOR mass spectrometry.

5. Fluorescence Light-Up Experiments

5.1 Preliminary Fluorescence Experiments at pH 5.5

We started the experiments with preliminary fluorescence experiments to see if the SPOC probe would bind to the bac rRNA A-site and if TO would intercalate into the dsRNA. pH 5.5 was chosen as it was reported that the tFIT probe required an acidic condition to form a triplex (Sato, Sato & Nishizawa, 2016).

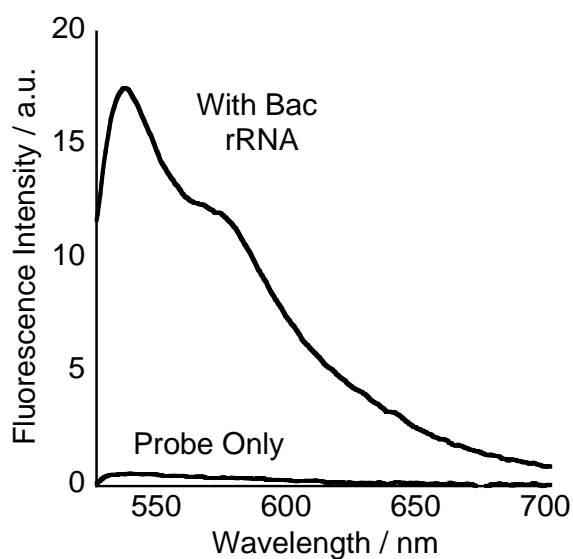


Figure 10. Fluorescence spectrum of the preliminary fluorescence experiments with SPOC ($n=2$) and Bac rRNA A-site. $[SPOC] = [RNA] = 50$ nM, pH 5.5, 25°C.

As seen in Fig. 10, the probe itself has negligible fluorescence. This is because when TO is unbound, it can lose energy through rotation around its monomethine bond, leading to non-emissive energy loss. However, upon the addition of Bac rRNA, there is a significant light-up response due to SPOC hybridising to the RNA and TO intercalating inside the dsRNA. TO's rotation is thus restricted and TO loses energy through emissive radiation in the form of fluorescence (Fig. 11). The SPOC probe is thus a fluorogenic probe (non-fluorescent when unbound, but fluorescent when bound) that exhibits a light-up response upon binding.

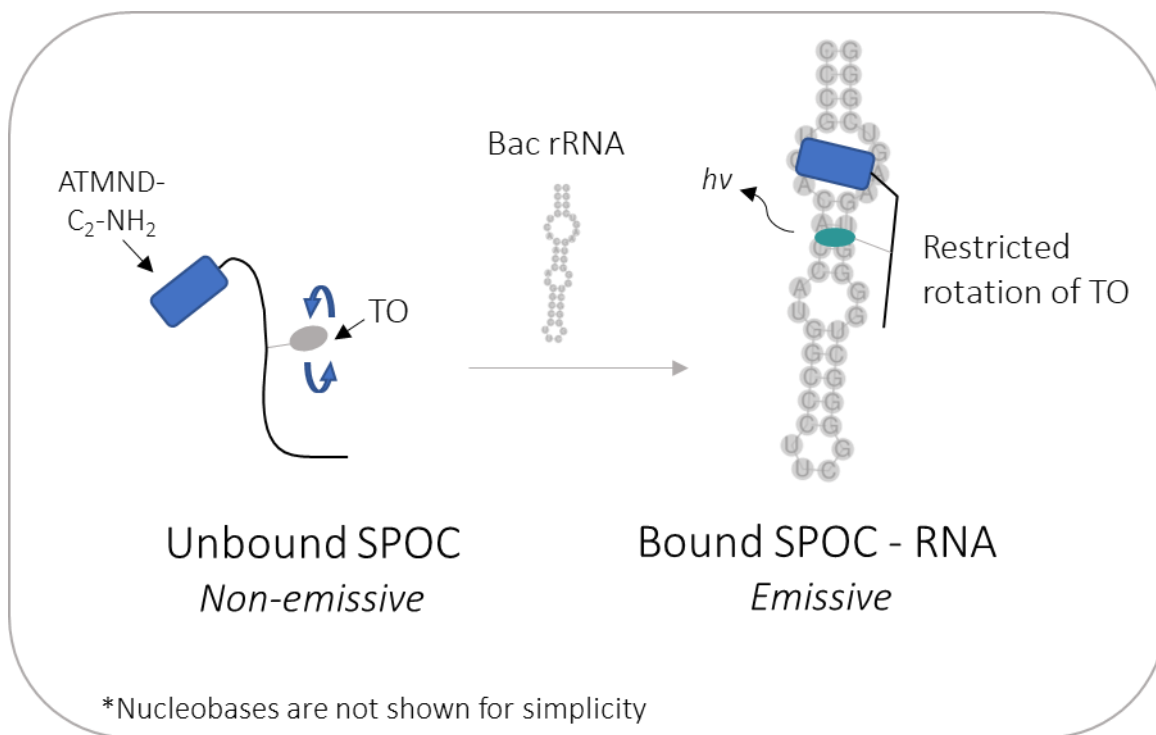


Figure 11. Illustration of why a light-up response is observed upon the addition of Bac rRNA to the SPOC probe.

5.2 Fluorescence Experiments at pH 7.0

It would be advantageous to use non-acidic pHs for the experiments because the human physiological pH is ~ 7.4 . In order to carry out FID assays that are close to the physiological conditions, we repeated the experiment at pH 7.0 and surprisingly, we still observed a light-up response, albeit at a much lower fluorescence intensity.

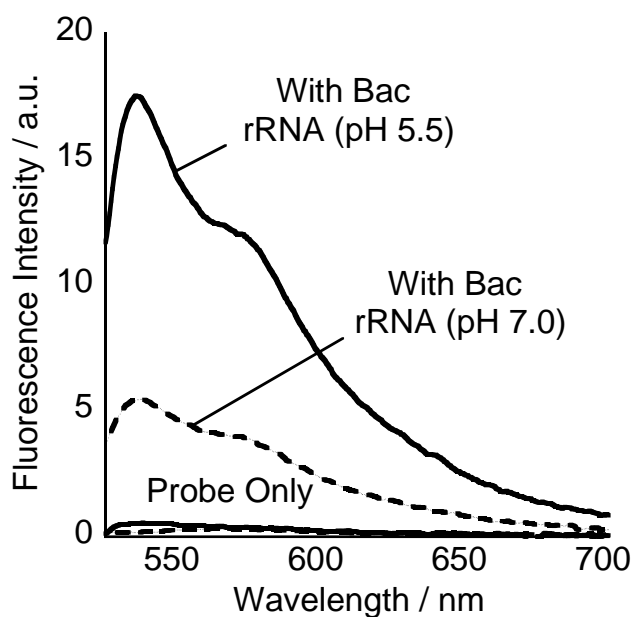



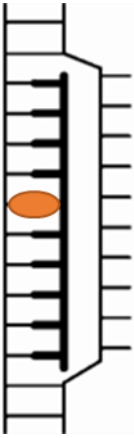
Figure 12. Fluorescence spectrum of the preliminary fluorescence experiments with SPOC ($n=2$) and Bac rRNA A-site. $[SPOC] = [RNA] = 50$ nM, pH 5.5 and pH 7.0, 25°C.

The weaker light-up response at pH 7.0 indicates that TO is intercalating into the dsRNA, implying that hybridisation of SPOC with Bac rRNA is still taking place, although to a much lesser degree compared to pH 5.5. This was unexpected because as explained previously, for Hoogsteen base-pairing to occur (triplex formation), cytosine has to be protonated. Since cytosine has a pK_a of ~ 4.4 , it requires an acidic environment for protonation, and thus a neutral pH would be extremely unfavourable. Furthermore, the SPOC probe sequence is almost entirely made up of cytosines, thus we assumed that SPOC binding would be very pH dependent.

5.3 Melting Point (T_m) experiments to determine Binding Mode

We wondered if SPOC was binding to Bac rRNA through duplex invasion that uses Watson-Crick base-pairing, thus circumventing the need for an acidic environment. Fluorescence experiments alone would not be able to discriminate between the two binding modes. One might argue that if the main binding mode was duplex invasion, then the fluorescence response at pH 5.5 and pH 7.0 should not differ significantly. However, at an acidic pH, one would still expect Hoogsteen base-pairing to take place and thus the combination of *both* triplex formation and duplex invasion could result in the higher fluorescence intensity observed at pH 5.5.

Table 2. Table comparing triplex formation and duplex invasion. The orange oval represents TO.

Binding Mode	Triplex Formation	Duplex Invasion
		
Type of Base-pairing	Hoogsteen	Watson-Crick
dsRNA : PNA	1 : 1	1 : 1
pH dependent	Yes	No

We therefore carried out melting point (T_m) experiments to see if triplex formation was still happening at pH 7.0. We measured the absorbance of the complexes at both 260 nm (wavelength at which nucleic acids absorb, data not shown) and 300 nm (wavelength that has the largest absorption difference between protonated and non-protonated cytosines) (Mergny, Lacroix, Han, Leroy & Helene, 1995).

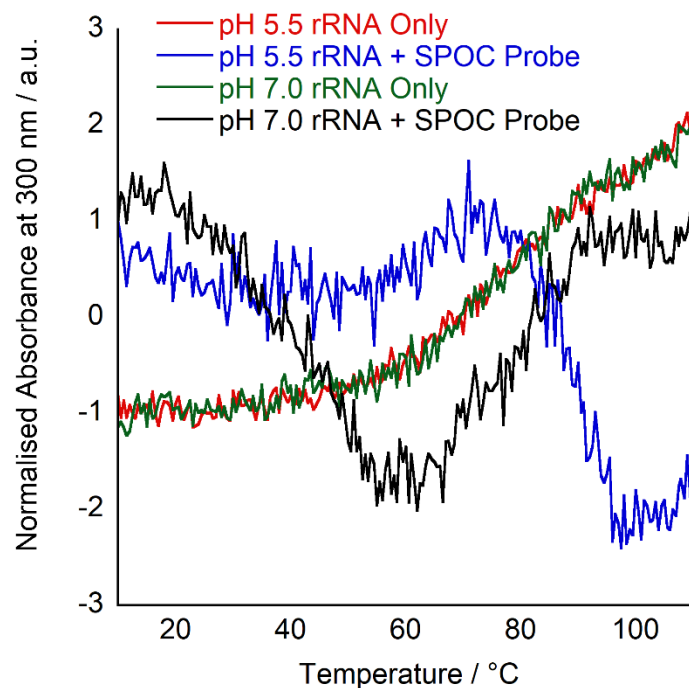


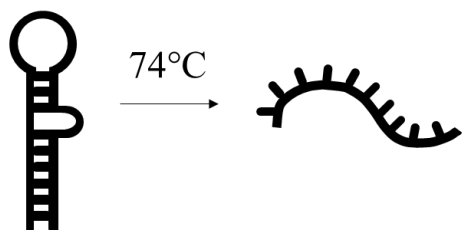
Figure 13. Melting curves of bacterial rRNA (3 μM) in the presence and absence of the SPOC probe (3 μM) (carrying ATMND-C₂-NH₂) at pH 5.5 or pH 7.0. Absorbance was measured at 300 nm.

Table 3. Table of the melting temperatures of the various samples at pH 5.5 or pH 7.0.				
	pH 5.5		pH 7.0	
Sample	rRNA Only	rRNA + SPOC	rRNA Only	rRNA + SPOC
T_m / °C	74	93	74	49, 74

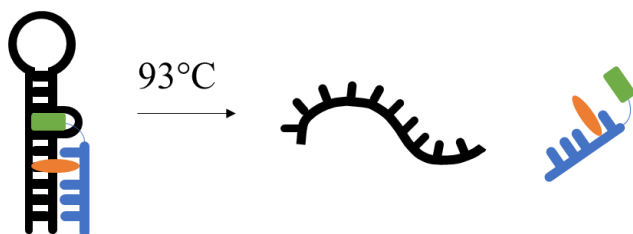
The melting temperature curve measured at 300 nm was chosen for the simplicity of explanation. From the Fig. 13, for both “rRNA Only” samples, a T_m of 74°C was observed. This corresponds to the dissociation of the dsRNA structure of the Bac rRNA hairpin into a single-stranded RNA. The identical T_m of the “rRNA only” samples show that the rRNA forms a dsRNA structure through pH independent Watson-Crick base-pairing.

For the “rRNA + SPOC” samples, at pH 5.5, a T_m of 93°C is observed, suggesting a single dissociation of the triplex into its individual strands. Comparatively, the sample at pH 7.0 was two T_m values, one 49°C and one at 74°C. The first T_m value corresponds to dissociation of the SPOC probe from the duplex RNA, and the second T_m value corresponds to the dissociation of the dsRNA hairpin into a single-stranded RNA, as per the “rRNA Only” samples. Thus, the T_m experiment shows that triplex formation is indeed happening even at pH 7.0.

rRNA Only (pH 5.5 & 7.0)



rRNA + SPOC (pH 5.5)



rRNA + SPOC (pH 7.0)

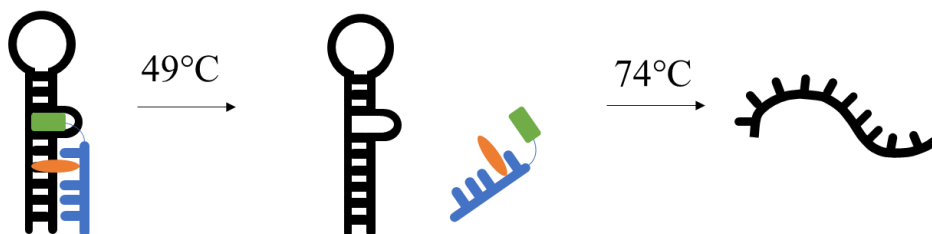


Figure 14. Illustration of the different melting points of the samples in the T_m experiment. The SPOC probe is represented by the blue figure, with TO in orange and ATMND-C₂-NH₂ in green.

5.4 Comparing the Fluorescence Response between the SPOC Probes

To understand how the different linker lengths of ATMND-C₂-NH₂, we compared the fluorescence response of the three different SPOC probes.

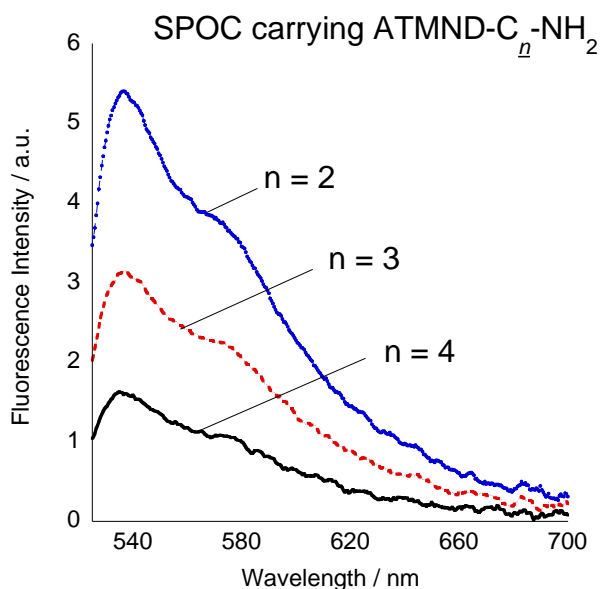


Figure 15. Fluorescence spectra of the three different SPOC probes ($n = 2 - 4$) with Bac rRNA at pH 7.0. $[\text{SPOC probe}] = [\text{Bac rRNA}] = 50 \text{ nM}$, 25°C .

The SPOC probe carrying ATMND-C₂-NH₂ results in the highest fluorescence response, followed by n=3 and n=4. This suggests that a longer and more flexible linker length does not facilitate TO intercalation into the dsRNA, thus resulting a weaker fluorescence response, however, the reason for that is still unknown.

As it is advantageous for an FID indicator to have as large a change in fluorescence as possible, we decided to do the rest of our experiments with the SPOC carrying ATMND-C₂-NH₂.

5.5 Reasons for Fluorescence Response at pH 7.0

As we sought to explain why SPOC can still form a triplex at pH 7.0, we uncovered two reasons – firstly, that ATMND- C_n -NH₂ acts as an anchor facilitating binding at pH 7.0; and secondly, that the flexible structure of the Bac rRNA itself facilitates binding at a neutral pH.

5.5.1 ATMND- C_n -NH₂ acts as an Anchor

In order to evaluate the contribution of ATMND- C_n -NH₂, we compared the SPOC probe with the control probe that has no ATMND moiety.

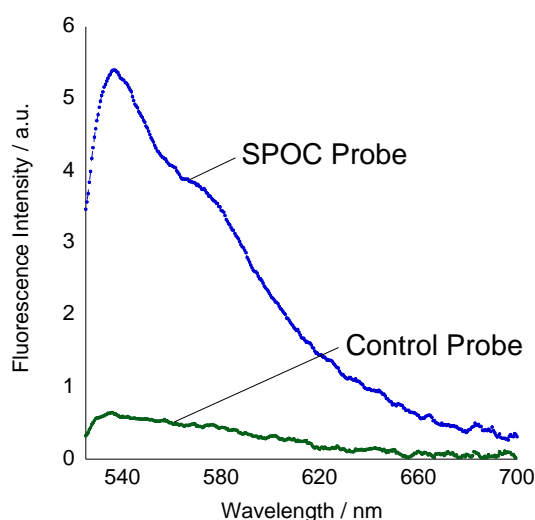


Figure 16. Fluorescence spectrum comparing the fluorescence response of the SPOC ($n=2$) probe and the control probe with Bac rRNA at pH 7.0. [Probe] = [Bac rRNA] = 50 nM.

As seen from Fig. 16, the fluorescence response of the control probe is much lower compared to the SPOC probe, showing how the conjugation of the ATMND moiety increases binding between the SPOC probe and the Bac rRNA.

To get a better idea of the degree that the ATMND moiety enhances binding affinity, the dissociation constant (K_d) value of SPOC-RNA complex can be measured and compared. The dissociation constant is a measure of the amount of dissociation that is happening. It follows that a lower K_d value means that less dissociation is happening, or in other words, that there is stronger binding between the two components.

Table 4. Table comparing K_d values of the SPOC probe (n=2) with the Control Probe at pH 5.5 and pH 7.0.

K_d / nM	SPOC Probe	Control Probe
pH 7.0	190 ± 72	3300 ± 240
pH 5.5	6.5 ± 3.4	33 ± 14

From Table 4, the SPOC probe has a lower K_d value compared to the control probe at both pHs, thus showing that the conjugation of the ATMND moiety increases the binding affinity between the SPOC probe and the dsRNA. The binding affinities at pH 5.5 are much higher compared to pH 7.0 for both probes, which is to be expected because Hoogsteen base-pairing is favoured at an acidic pH. However, as the pH increases from pH 5.5 to pH 7.0, the SPOC probe's K_d value increases by ~30 times, while the control probe's K_d value increases by 100-fold. We can conclude that the SPOC probe is therefore more pH independent compared to the control probe, and that the pH independence is due to the presence of the ATMND moiety.

As such, as the conjugation of the ATMND moiety not only increases the fluorescence intensity of the SPOC probe, but also strengthens the binding affinity and ascribes pH independence to the probe, we have chosen to describe ATMND's contribution as it acting as an anchor.

5.4.2 Flexible Bac rRNA Structure facilitates Binding

The previous report of the tFIT probe described how the tFIT probe could not bind at pH 7.0 (Sato, Sato & Nishizawa, 2016). However, the control probe, which is essentially a tFIT probe, still shows a small response to the Bac rRNA even at pH 7.0 (Fig. 16). The RNA used in the previous report was a fully-matched hairpin RNA. In contrast, the Bac rRNA A-site contains an internal loop as well as several mismatched base-pairs. We wondered if the flexible structure of the Bac rRNA A-site itself was contributing to the binding at neutral pH.

To test this hypothesis, we used a mutant Bac rRNA that had fully-matched base-pairs except for the internal loop region.

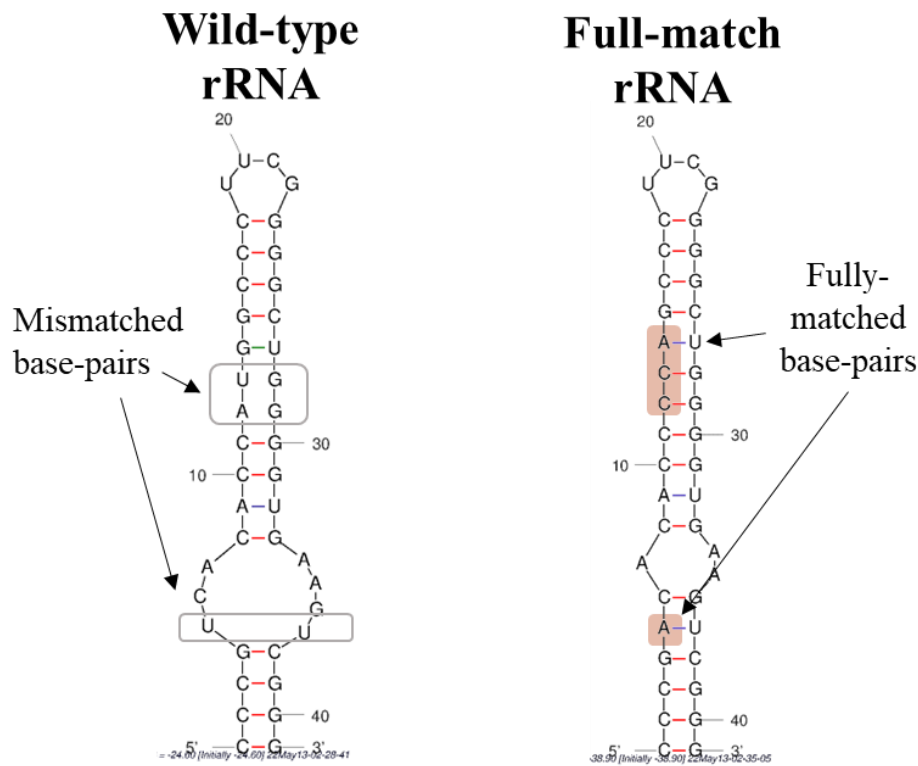


Figure 17. RNA sequence of the wild-type Bac rRNA A-site and the mutant full-match rRNA A-site. The secondary structures have been predicted by mfold (Accessed through: <http://www.unafold.org/mfold/applications/rna-folding-form.php>).

The predicted secondary structures show that the full-match rRNA has a much more rigid structure compared to the wild-type rRNA.

Fluorescence experiments of SPOC and the two different rRNAs reveal that SPOC binds to the wild-type rRNA with a much stronger fluorescence intensity (~7-fold higher, 537 nm) (Fig. 18). Moreover, SPOC binds to the full-match rRNA with a K_d value that is one order of magnitude higher compared to the wild-type rRNA. This shows that the flexible structure of the wild-type Bac rRNA itself contributes to the ability of SPOC to bind even at a neutral pH. While serendipitous, many wild-type RNAs of significance contain mismatched base-pairs (eg. Influenza A virus promoter region (Fig. 22), Japanese encephalitis virus 3' untranslated region (Chen et al., 2018); Moloney Murine Leukemia Virus Ψ -site (D'Souza et al., 2004), etc.) As such, we can expect that the SPOC probe design would also work at neutral pHs for other dsRNA targets.

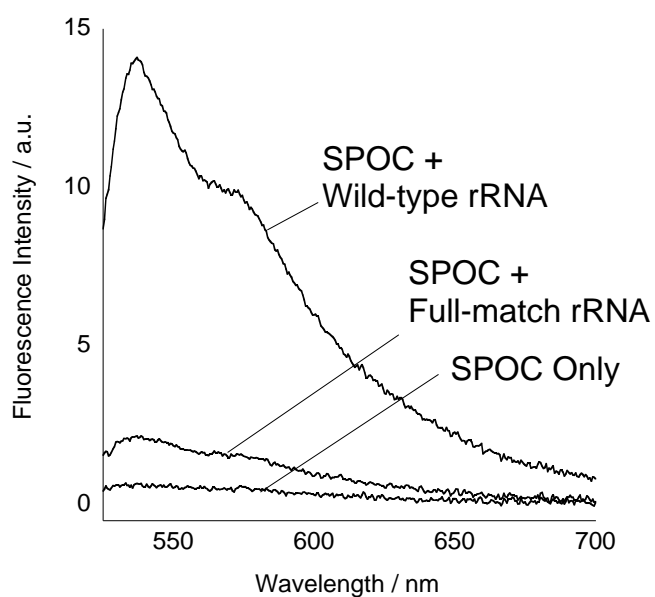


Figure 18. Fluorescence spectra of SPOC with the wild-type rRNA and full-match rRNA. [SPOC] = [rRNA] = 100 nM, pH 7.0, 25°C.

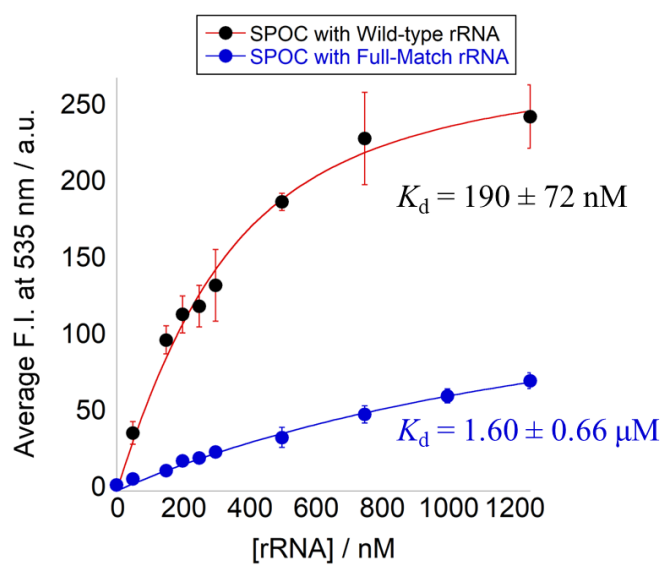


Figure 19. Concentration dependence curves (annotated with measured K_d values) of SPOC with wild-type and full-match rRNAs. [SPOC] = 250 nM, [rRNA] = 0 – 1250 nM, pH 7.0, 25°C. Error bars were plotted from three independent experiments.

6. Binding Affinity Experiments

The binding affinities of the various probes were examined by measuring their K_d values when binding with Bac rRNA A-site.

The K_d value of the SPOC probe was calculated to be 190 ± 72 nM, which was much lower compared to its two components – the tFIT probe (control probe) (3300 ± 240 nM) and ATMND- C_2 -NH₂ (7800 ± 610 nM). This shows that conjugation is a good strategy to increase the binding affinity of a probe to a dsRNA. Furthermore, ATMND- C_2 -NH₂ was previously reported to be the tightest binding non-aminoglycosidic ligand to the Bac rRNA A-site. Using the conjugation strategy, SPOC is now the tightest non-aminoglycosidic binder to the Bac rRNA A-site.

Moreover, Rozner and his group developed a 9-mer probe for the Bac rRNA A-site, where they sought to overcome the pH dependence by using an artificial nucleobase, E (Gupta, Zengeya & Rozners, 2011). The K_d value of that probe was 190 nM (pH 6.25) as well. This shows that a simple conjugation strategy can produce probe that rivals even more sophisticated probe designs.

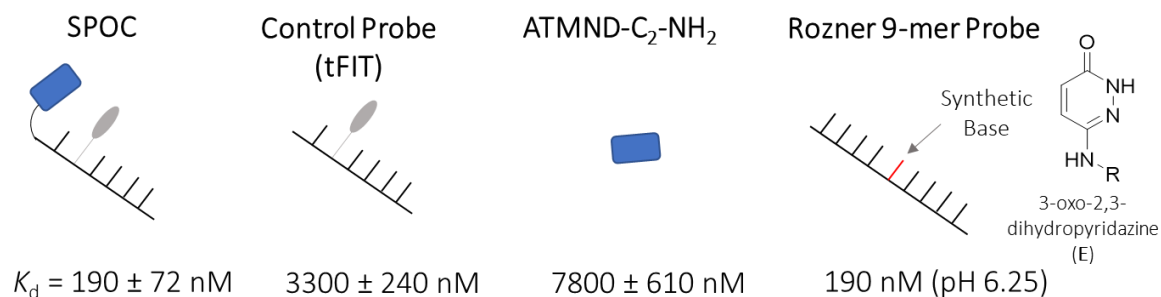


Figure 20. Illustrative summary of the K_d values of SPOC, its individual components, and Rozner's 9-mer probe.

The binding affinities of SPOC with varying linker lengths of ATMND- C_n -NH₂ were also measured. Surprisingly, the binding affinities of all three probes were not significantly different from each other, although the fluorescence intensity of the SPOC (n=2) probe was much larger compared to the other two probes (Fig. 21).

What this seems to suggest is that although the varying linker lengths of the ATMND moiety affects how well TO intercalates into the dsRNA (thereby affecting the fluorescence intensity), the three versions of the ATMND moiety increase the binding of the SPOC probe to similar extents. However, as previously stated, in order to maximise the change in fluorescence as an FID indicator, SPOC (n=2) was chosen to for use in subsequent experiments.

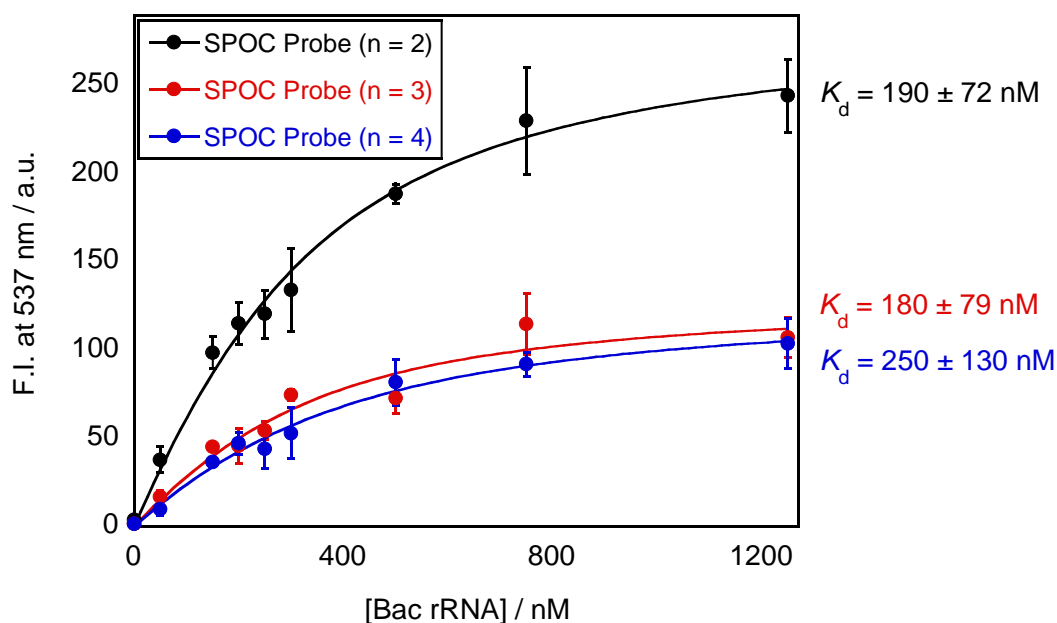


Figure 21. Combined concentration dependence curve of SPOC (n=2 – 4) with Bac rRNA A-site at pH 7.0. [SPOC] = 250 nM, [Bac rRNA] = 0 -1250 nM, 25°C.

7. Selectivity Experiments

Selectivity experiments were carried out by comparing SPOC's response to four different RNAs –bac rRNA A-site (target), human cytoplasmic rRNA A-site (1-mismatch), human mitochondrial rRNA A-site (1 mismatch) and an unrelated RNA, influenza A viral RNA (vRNA) promoter region.

As seen from Fig. 23, SPOC exhibits the highest fluorescence response with the target RNA (Bac rRNA) by having a fluorescence intensity at least a three-fold higher compared to the other RNAs. Moreover, the K_d values of SPOC to the non-target human rRNA A-sites are one order of magnitude higher compared of that to Bac rRNA A-site (Fig. 24). This shows that the SPOC probe is very selective to the target RNA over the non-targets, and that RNA sequences that contain just one mismatch with the SPOC probe result in a significant decrease in binding affinity and fluorescence intensity. This sequence specificity is the result of the conjugation with a tFIT probe that sequence-specifically binds to the dsRNA. This selectivity towards the target RNA would prove useful for FID indicators as this would prevent the indicator from binding to undesired regions of the target dsRNA, decreasing the probability for false positives / negatives.

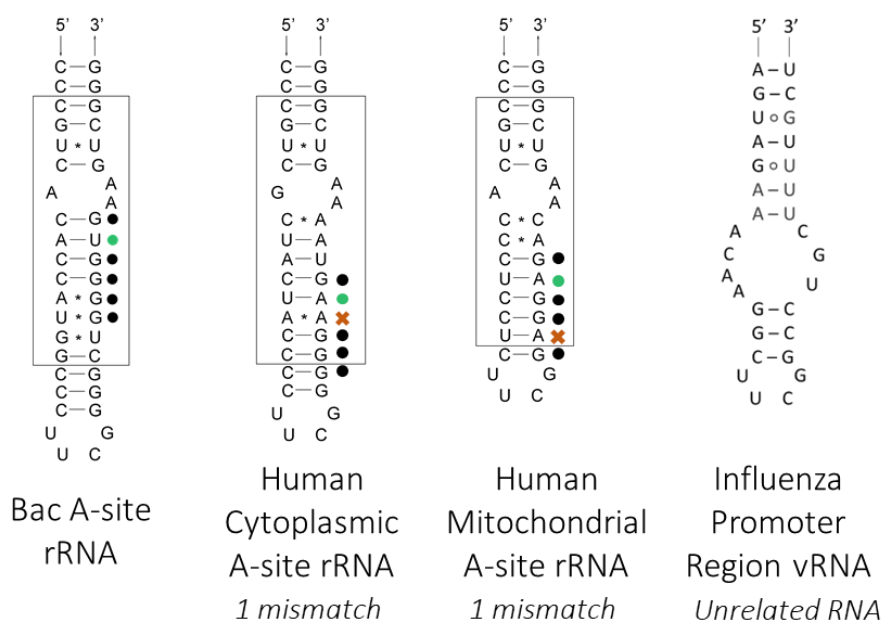


Figure 22. RNA sequences of the four RNAs used in the selectivity experiments. Boxed regions represent the native sequence of the RNAs. SPOC's expected binding region has been indicated by dots. TO is represented by a green dot, while a mismatch is represented by an X.

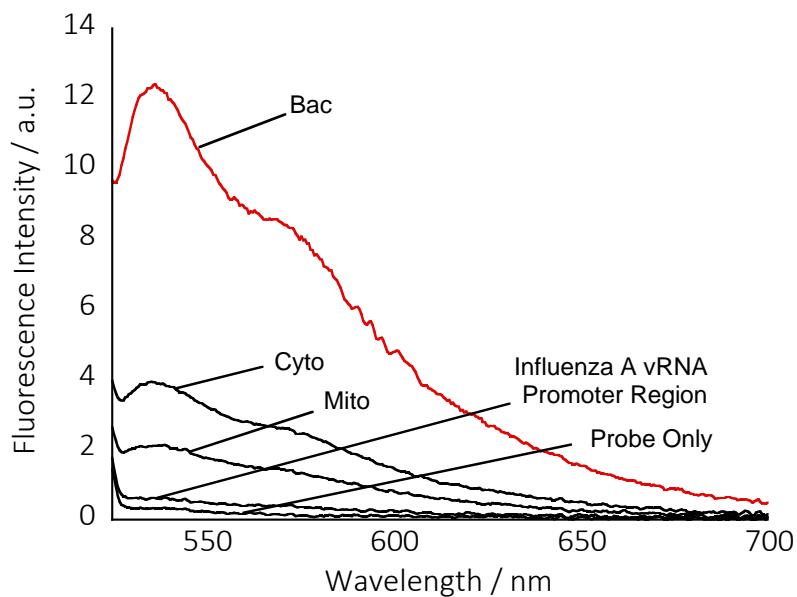


Figure 23. Fluorescence spectrum of SPOC with four different RNAs. [SPOC] = [RNA] = 50 nM, pH 7.0, 25°C.

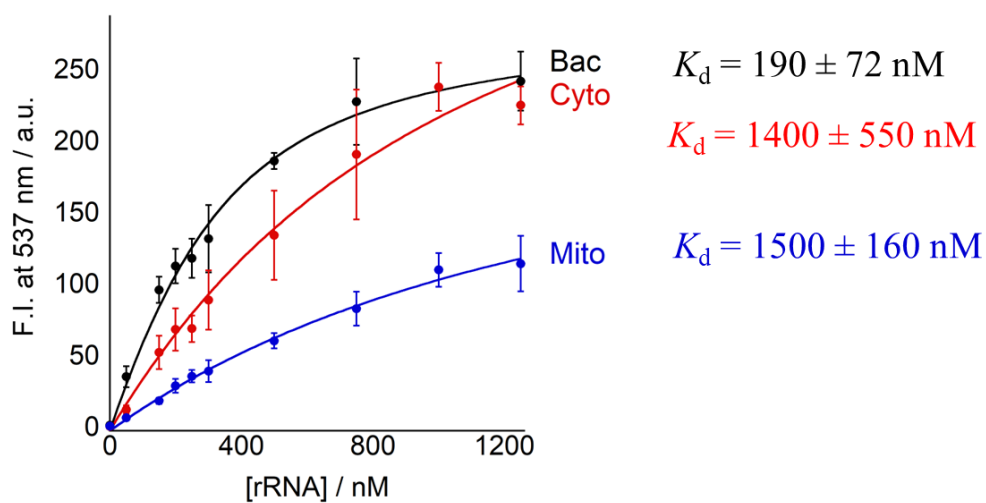
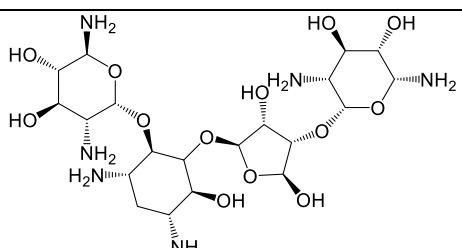
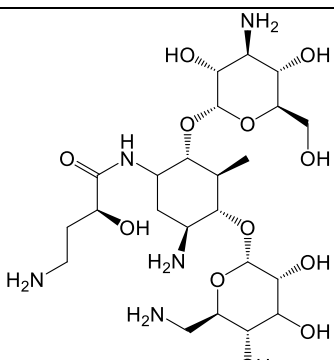
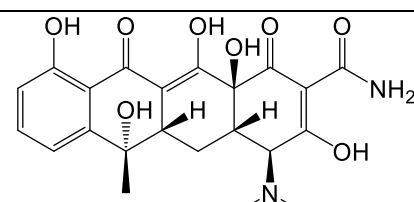


Figure 24. Combined concentration dependence curves of SPOC with three different rRNA A-sites. The corresponding K_d values have been added. [SPOC] = 250 nM, [rRNA] = 0 – 1250 nM, pH 7.0, 25°C. The error bars were plotted from three independent experiments.

8. Mock FID Assay

Finally, a mock FID assay was carried out to assess SPOC's potential as an FID indicator. We used test compounds whose binding affinity with Bac rRNA A-site had previously been evaluated, hence, they are not true test compounds, in the sense that they are not new test compounds. Therefore, the FID assay conducted was a mock FID assay.

The test compounds chosen for this mock FID assay were neomycin, amikacin and tetracycline. Neomycin and amikacin are aminoglycosides, and their K_d values with Bac rRNA A-site have been reported. Tetracycline is not known to bind to the bac A-site, and serves as a negative control. Neomycin has the lowest K_d value (0.21 nM) (Kaul & Pilch, 2002), followed by amikacin ($K_d = 8.1 \mu\text{M}$) (Dudek, Romanowska, Witula & Trylska, 2014). Hence, we would expect neomycin to result in the most amount of displacement, and for amikacin to result in some displacement.

Test Compound	Chemical Structure	K_d
Neomycin		0.21 nM
Amikacin		8.1 μM
Tetracycline		N.A.

From Fig. 25, without any test compound, SPOC has a fluorescent intensity of around 72 a.u. The addition of tetracycline causes negligible change. However, the addition of amikacin results in a decrease of F.I. to

43 a.u. and the addition of neomycin causes a decrease of F.I. to 8 a.u. The degree of displacement is in good correlation with the reported K_d values. Therefore, SPOC shows potential for use as an FID indicator.

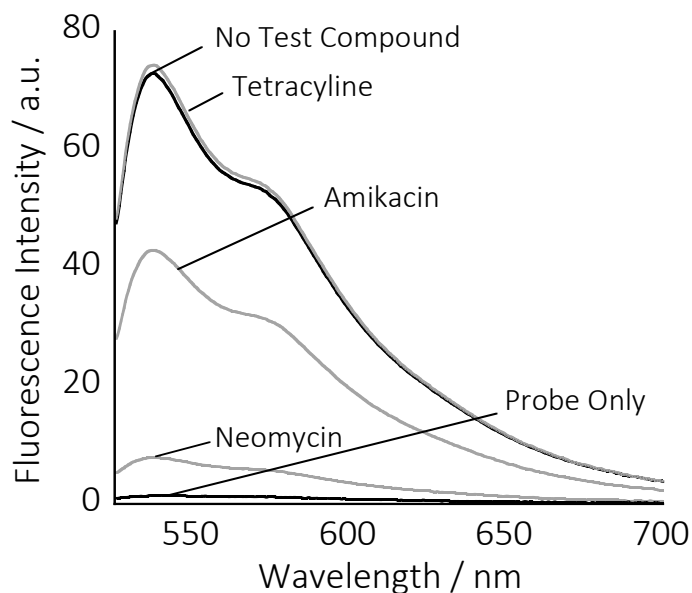


Figure 25. Fluorescence spectrum of the mock FID assay with SPOC and three test compounds. [SPOC] = 1 μ M, [Test Compound] = 5 μ M, pH 7.0, 25°C.

8.1 Investigating the effects of a Lower Concentration

It is desirable to minimise the amount of test compounds needed in an FID assay, as some test compounds could be extremely valuable and expensive. Being able to use smaller amounts of test compounds thus decreases the overall cost of the assay, making the assay more accessible to laboratories with smaller budgets as well.

To investigate if SPOC could still work at a lower concentration, we decreased the concentration of SPOC from 1 μ M to 100 nM. Under these more stringent conditions, SPOC was still able to be displaced by neomycin, but not by amikacin anymore. This shows that just by varying the experimental conditions, SPOC could be used to as an indicator that can sieve out moderate to strong binders (1 μ M), or to only sieve out strong binders (100 nM), broadening the scope of SPOC's usefulness.

In contrast, at 100 nM, ATMND-C₂-NH₂, which had previously been used as an FID indicator (Sato et al., 2018), cannot discriminate between test compounds at all. This clearly shows how SPOC's discriminatory ability is a superior FID indicator compared to ATMND-C₂-NH₂.

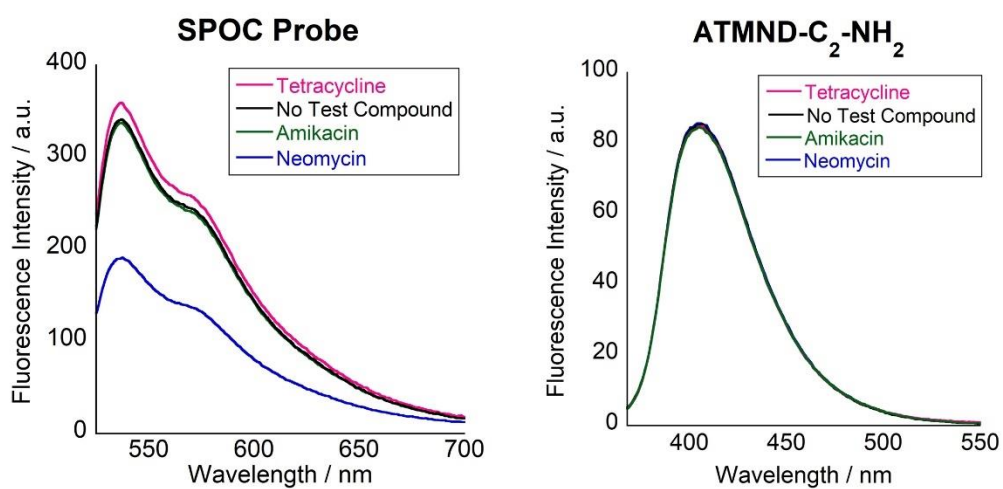


Figure 26. Fluorescence response of the SPOC probe (100 nM) and ATMND-C₂-NH₂ (100 nM) with Bac rRNA A-site (100 nM) in the presence and absence of test compounds (500 nM) at 25°C, pH 7.0. Ex wavelength: 515 nm (SPOC probe), 358 nm (ATMND-C₂-NH₂).

8.2 Measuring the Limit of Detection of Neomycin

As neomycin (500 nM) was still able to displace SPOC (100 nM), we determined the limit of detection (LOD) of neomycin when [SPOC] = 100 nM.

We calculated the LOD of neomycin using the following equation:

$$LOD = 3\left(\frac{S_{blank}}{Slope}\right)$$

where S_{blank} is the standard deviation of the blank, and the *slope* refers to the slope of the calibration curve. The absolute value of the calculated slope was 23.1, and the standard deviation of the blank was 0.327.

As such, the calculated LOD of neomycin is 42 nM.

These results tell us that an even lower amount of neomycin (< 500 nM) could still displace SPOC. Thus, the required amount of test compounds could be reduced even more for strong binders to the A-site, further decreasing the cost of the FID assay.

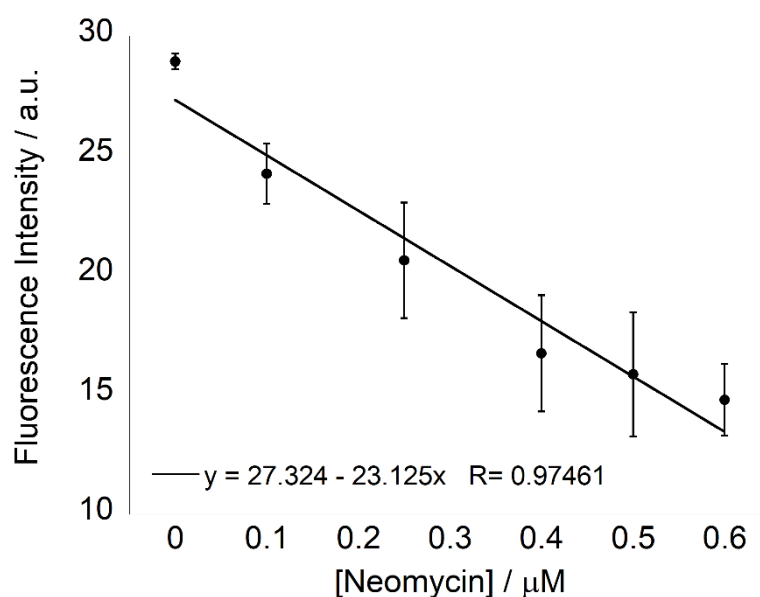


Figure 27. Calibration curve for SPOC (0.1 μM) and Bac rRNA A-site (0.1 μM) upon the addition of neomycin (0 – 0.6 μM), pH 7.0, 25°C. Error bars were plotted from three independent experiments.

9. Conclusion and New Generations of SPOC

In conclusion, the simple conjugation strategy of a small molecule with a PNA oligomer (tFIT probe) resulted in a probe that has higher binding affinity to the Bac rRNA A-site, that is selective to the target RNA and that shows potential for use in an FID assay. We have shown that ATMND-C₂-NH₂ acts as an anchor for SPOC, allowing SPOC to bind even at a neutral pH, something that other dsRNA-targeting PNA probes have yet to be able to achieve (Gupta, Zengeya & Rozners, 2011; Sato, Sato & Nishizawa, 2016). Moreover, the SPOC probe is now the tightest-binding non-aminoglycosidic ligand to the Bac A-site rRNA, with a K_d of 190 nM.

The rational design of SPOC allows for it to be tailor-made to fit other dsRNA targets. In fact, in our laboratory, a similar small molecule – PNA oligomer conjugate probe has been designed and evaluated for the influenza A virus (IAV) promoter region RNA (Sato et al., 2022). The small molecule chosen was DPQ, which has been reported to bind to the influenza promoter region internal loop ($K_d = 50.5 \mu\text{M}$) (Lee et al., 2014).

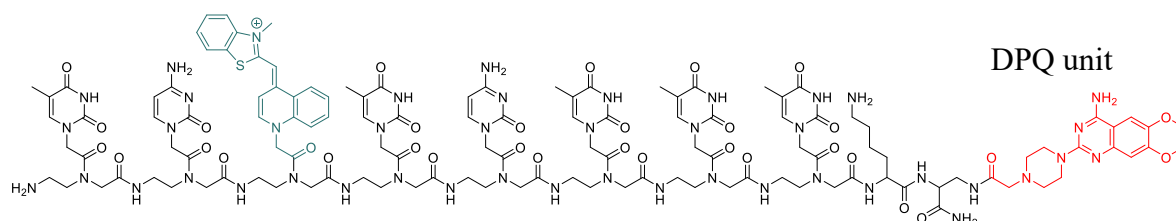


Figure 28. Chemical structure of the SPOC probe for the influenza A virus promoter region. Probe sequence: NH₂-TC(TO)TCTTT-Lys-Dap-DPQ. TO has been highlighted in teal, while DPQ has been highlighted in red.

When DPQ was conjugated to a tFIT probe, the resultant conjugate probe was similarly able to bind to the IAV RNA at pH 7.0, with a low K_d value of 107 nM. It was also used as an FID indicator in a mock FID assay with neomycin, paramomycin and DPQ itself, as well as a viral inhibitor itself (results yet to be published).

Furthermore, we are looking into small molecule alternatives by investigating if additional nucleobases could be a substitute as an internal loop binder instead of a small molecule (Han, 2022). This 9-mer probe showed optimistic results that additional nucleobases can also work as an internal loop binder, and also showed fluorescence response at a neutral pH.

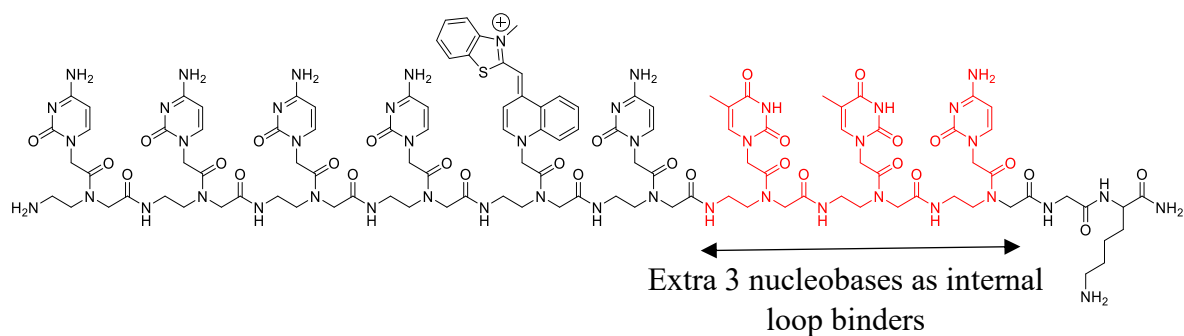


Figure 29. Chemical structure of 9-mer probe for the Bac rRNA A-site, where an additional three nucleobases (highlighted in red) have been used instead of ATMND-C₂-NH₂ to bind to the internal loop.

However, as simple and effective as the conjugation strategy is, it faces its fair share of limitations. For example, for Hoogsteen base-pairing to work, the dsRNA target must contain a stretch of purine nucleobases for the tFIT oligomer to base-pair with. Moreover, SPOC requires a suitable small molecule that can bind to the target dsRNA. The target dsRNA would also preferably contain an internal loop / bulge. If these requirements are not met, the SPOC strategy cannot be effectively applied.

As such, other new and creative dsRNA-targeting probe designs are still necessary, and we explore another probe design in **Chapter 3**.

10. Experimental Methods

10.1 General Information

Rink Amide Resin was purchased from Merck Millipore (Billerica, MA). Fmoc-protected amino acids were purchased from Sigma-Aldrich (Darmstadt, Germany) and Watanabe Chemical Industries (Hiroshima, Japan). Fmoc/Bhoc protected PNA monomers were purchased from Panagene (Daejeon, South Korea). Other reagents were commercially available and of analytical grade and were used without further purification. Model hairpin rRNAs were custom synthesized and HPLC purified (99%) by GeneDesign, Inc. (Osaka, Japan). The concentration of the RNAs was determined from the OD₂₆₀ at 25°C and calculated using the OD₂₆₀ = 1 = 40 µg/mL. Ultrapure water (18.2 MΩ-cm specific resistance) was used from an Elix 5 UV water purification system and a MilliQ Synthesis A10 system (Millipore Co., Bedford, MA), followed by filtration through a BioPak filter (Millipore Co.) in order to remove RNase.

Unless otherwise mentioned, all measurements were performed in 10 mM sodium acetate buffer solutions (pH 5.5) or sodium phosphate buffer solutions (pH 7.0) containing 100 mM NaCl and 1.0 mM EDTA. Before measurements, annealing of the RNA-containing samples was carried out as follows: heating at 95°C for 10 min, before cooling by 1°C for 1 min all the way down to 25°C, where it is held until the next step.

10.2 Synthesis of ATMND-C₂-NH₂

ATMND-C₃-NH₂ and ATMND-C₄-NH₂ were available in the laboratory and did not have to be resynthesised. Only ATMND-C₂-NH₂ had to be synthesised and was characterised by ¹H NMR and ESI-mass spectrometry.

¹H NMR (500 MHz, CDCl₃): δ (ppm) = 7.98 (d, 1H, *J* = 8.5 Hz), 6.64 (d, 1H, *J* = 9 Hz), 3.72 (q, 2H, *J* = 5.5 Hz), 3.03 (t, 2H, *J* = 5.5 Hz), 2.69 (s, 3H), 2.51 (s, 3H), 2.34 (s, 3H).

10.3 Synthesis of Probe and Control Probe

Preparation: An appropriate amount of Rink Amide Resin was measured out (usually 10 µmole) and left to swell in dehydrated dichloromethane (DCM) overnight. DCM, dimethylformamide (DMF) and N-methylpyrrolidinone (NMP) were dehydrated with molecular sieves (4 Å) for at least one night. DCM, DMF and NMP were purchased from FUJIFILM Wako Pure Chemical Corporation, peptide synthesis grade. As far as possible, PNA synthesis was done with help from the Biotage® Initiator+ (Biotage, Uppsala, Sweden) inside a corresponding microwave peptide vial.

Fmoc Deprotection: In order to remove the fmoc-protecting groups on the resin, the resin was stirred with 20% piperidine (in DMF) (FUJIFILM Wako Pure Chemical Corporation, peptide synthesis grade) for one minute, before adding an additional 100 μ L of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) and stirring for another minute. The liquids were drained, and this deprotection step was repeated for another two rounds. After draining, the resin was washed with two rounds of DMF (1 mL, 1 min per round), then two rounds of DCM (1 mL, 1 min per round), then four rounds of DMF (1 mL, 1 min per round). A Kaiser test would be done at this point to check if the deprotection had taken place. If the Kaiser test was unsuccessful, another round of deprotection would be done. If successful, the coupling step would commence.

Coupling of monomers: The PNA monomer / amino acid (4 eq. wrt resin) and a coupling agent, COMU (3.9 eq.) (Wako Chemicals, Organic synthesis grade), were dissolved in 800 μ L of DMF before 12 eq. of diisopropyl-ethylamine (DIEA) (Tokyo Chemical Industry Co. Ltd.) was added. This solution was transferred to the microwave vial and topped up to 2 mL before inserting into the Biotage[®] Initiator+. The following conditions were used: Temperature: 75°C, Time: 10 min, Pre-stirring: 15 sec, Pressure: Off, Absorbance: High/ Very High, Vial type: 2.0-5.0 mL, Fixed Hold Time: On. The coupling step was done twice, after which the resin would be washed twice with DMF, twice with DCM and four times with DMF. Another Kaiser test would be done to see if all the active groups had been coupled. If unsuccessful, another round of coupling would be done. If successful, the capping step would commence.

Capping: A capping solution of Acetic Anhydride: 2,6-Lutidine: DMF (5: 6: 89) was made the resin would be stirred with 1 mL of the capping solution for 2 min. Acetic Anhydride and 2,6-Lutidine were both purchased from FUJIFILM Wako Pure Chemical Corporation. This step was repeated twice.

Alloc deprotection: 8 eq. of borane dimethyl amine complex (DMAB) (Sigma-Aldrich, 97%) and 1 eq. of tetrakis-triphenylphosphine palladium ($\text{Pd}(\text{PPh}_3)_4$) (Tokyo Chemical Industry Co. Ltd.) were dissolved in 1 mL of DCM each. The DMAB solution would be added to the resin and vortexed for one minute before adding the $\text{Pd}(\text{PPh}_3)_4$ solution. The combined mixture would be shaken for 45 min. The liquids are drained and the resin is washed until the filtrate turns colourless. Another round of the deprotection step would be carried out. A chloranil test (Kanto Chemical Co. Inc.) would be done to check if the alloc deprotection had succeeded.

Coupling of fluorophore: 4 eq. of the fluorophore, 3.9 eq. of COMU and 4 eq. of pyridium p-sulfonate (PPTS) (Sigma-Aldrich, 96%) were dissolved in 1 mL of DMF. 12 eq. of DIEA was added and the resultant

solution was added to the resin and vortexed overnight. Subsequently, the liquids would be drained and the resin would be wash with DMF, DCM and DMSO until the filtrate turns colourless. Generally, the fluorophore coupling step would be repeated two more times.

Bhoc / Boc deprotection and cleaving from the resin: A TFA: m-cresol (85:15) solution was prepared. M-cresol was purchased from Tokyo Chemical Industry Co. Ltd. The solution would be added to the resin until it just covers the top of the resin and the resultant mixture would be shaken for a total of 3 hours, with 2 min of vortexing ever half an hour. The filtrate would be collected in PE centrifuge tubes (GE Healthcare UK Limited, Buckinghamshire, UK) and topped up to 10 mL with cold diethyl ether to precipitate the probes. The tubes would be left in the freezer overnight and the probes would be collected by centrifugation the following day.

Purification: The crude probes were purified via HPLC with an Inertsil ODS - 3 5 μm (20 x 250 mm) column (GL Sciences, Torrance, CA), with JASCO UV 2070-Plus spectrometer (Japan Spectroscopic Co. Ltd., Tokyo, Japan). The wavelengths were set to 510 nm (for TO) and 260 nm (for PNA). The oven temperature was set to 55°C. The solvents used were 0.1% TFA H₂O and 0.1% TFA acetonitrile (Sigma-Aldrich, HPLC grade). Typically, the solvent gradient would be H₂O : Acetonitrile 90:10 \rightarrow 50: 50. The separated fractions would be checked using MALDI-TOF mass spectrometry (matrix: CHCA) (Bruker Daltonics autoflex Speed-S1, Bruker, Billerica, MA) to confirm the presence of the purified probe. The concentration of the control probe was calculated in water via UV-vis spectroscopy at 25°C using the following molar extinction coefficients at 260 nm: 8800 M⁻¹cm⁻¹ for thymine, 7300 M⁻¹cm⁻¹ for cytosine, and 9400 M⁻¹ cm⁻¹ for TO. For the SPOC probes, the following molar coefficients of the ATMND moieties were used.

Table 6. Table of molar extinction coefficients used to calculate the SPOC probes' concentrations.

ATMND moiety	$\epsilon / \text{cm}^{-1} \text{M}^{-1}$
ATMND-C ₂ -NH ₂	15 187
ATMND-C ₃ -NH ₂	14 587
ATMND-C ₄ -NH ₂	14 625

The mass spectra, HPLC chromatograms and UV-Vis spectra of the all the probes are available in the supporting information (Fig. S1 – S4).

10.4 Spectroscopic Experiments

10.4.1 UV-Vis Spectroscopic Experiments

UV-Vis experiments were done with a JASCO model V-570 UV-Vis spectrophotometer (Japan Spectroscopic Co. Ltd., Tokyo, Japan) at 25°C. Two 2 × 10 mm quartz cells (optical path length: 10 mm) were used – one for the sample, and the other for the reference (ultrapure water). A baseline correction was first done using only the buffer before UV measurement of the samples.

10.4.2 General Fluorescence Experiments

All fluorescence experiments were carried out using a JASCO FP-6500 spectrofluorophotometer (Japan Spectroscopic Co. Ltd., Tokyo, Japan), with a 3 x 3 mm quartz cell at 25°C. Excitation and emission band widths were set to 5 nm, and the sensitivity was set to medium, or low in the case of the mock FID assay. An excitation wavelength of 513 nm was used for TO and 358.5 nm for ATMND-C₂-NH₂.

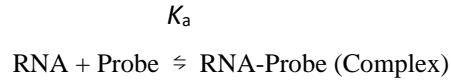
10.4.3 Fluorescence Titration Experiments

Fluorescence titration experiments were carried out in order to calculate the dissociation constants. The concentration of the probe was held constant while the concentration of the rRNA was varied (up to 5 eq. of the probe concentration). Excitation and emission band widths were set to 5 nm, and the sensitivity was set to medium. An excitation wavelength of 515 nm (n=2), 513 nm (n=3) and 518 nm (n=4) was used for SPOC probes and 358.5 nm for ATMND-C₂-NH₂. The fluorescence peak values for each concentration of rRNA was used to plot a graph and a 1:1 binding isotherm was fitted to it in order to evaluate the association constant (K_a). The dissociation constant is the reciprocal of K_a .

10.5 Determining the Curve-Fitting Equations

The curve-fitting equations used were:

For TO,



The association constant can be represented using the equilibrium concentrations of the RNA, Probe and RNA-Probe complex as such:

$$K_a = \frac{[\text{Complex}]}{[\text{RNA}][\text{Probe}]}$$

The mass balance equations for the RNA and the Probe would be:

$$[\text{RNA}]_0 = [\text{RNA}] + [\text{Complex}]$$
$$[\text{Probe}]_0 = [\text{Probe}] + [\text{Complex}]$$

Using the above equations, we can express [Complex] using only the initial concentrations of the experiment:

$$[\text{Complex}]^2 - [\text{Complex}] \left([\text{RNA}]_0 + [\text{Probe}]_0 + \frac{1}{K_a} \right) - [\text{RNA}]_0 [\text{Probe}]_0 = 0$$
$$[\text{Complex}] = \frac{1}{2} \left([\text{Complex}] \left([\text{RNA}]_0 + [\text{Probe}]_0 + \frac{1}{K_a} \right) - \sqrt{\left([\text{RNA}]_0 + [\text{Probe}]_0 + \frac{1}{K_a} \right)^2 + 4[\text{RNA}]_0 [\text{Probe}]_0} \right)$$

which is Equation 1.

The change in concentration of the RNA-Probe complex can be monitored from the change in a physical property (F), which in this case is its fluorescence intensity. F , the raw fluorescence, is normally a sum of the fluorescence of all the various components, and thus can be expressed in the following way:

$$F = F_{Probe}[Probe] + F_{RNA}[RNA] + F_{RNA-Probe}[RNA - Probe]$$

By assuming that free RNA does not fluoresce, we can simplify the equation to be:

$$F = F_{Probe} + \left(\frac{[Complex]}{[Probe]_0} \right) (F_{Complex} - F_{Probe})$$

$$\Delta F = F_{\Delta Complex} \left(\frac{[Complex]}{[Probe]_0} \right)$$

By substituting in Equation 1, ultimately, the following equation was used to fit the obtained curves:

$$\Delta F = \frac{F_{\Delta HG}}{2[H]_0} \left([G]_0 + [H]_0 + \frac{1}{K_a} - \sqrt{\left([G]_0 + [H]_0 + \frac{1}{K_a} \right)^2 - 4[H]_0[G]_0} \right)$$

where H represents the probe, G presents the RNA and HG represents the complex.

As mentioned previously, the dissociation constant is the reciprocal of K_a .

For ATMND-C₂-NH₂,

Normally, the curve-fitting equation used for ATMND-C₂-NH₂ would be:

$$\frac{F}{F_0} = \frac{1 + \frac{k_{complex}}{k_{RNA}} K_a [RNA]}{1 + K_a [RNA]}$$

Where $k_{complex}$ and k_{RNA} are the proportionality constants of the complex and the RNA respectively.

Unfortunately for ATMND-C₂-NH₂, the curves obtained did not appear to have reached saturation, and thus it was unclear if the $k_{complex}/k_{RNA}$ term was needed or not. Eventually, a much simpler equation was used instead:

$$\frac{F}{F_0} = 1 - \frac{[Complex]}{[Probe]_0}$$

This equation assumes that the fluorescence of ATMND-C₂-NH₂ is quenched as more and more complex is formed. It also assumes that the complex does not fluoresce.

10.6 Fluorescent Indicator Displacement Assay Experiment

All the conditions used were the same as those of the fluorescence spectroscopic experiments. The test compounds were added in after the RNA had been reannealed and resulting samples were incubated for 30 min. The probe was added in just 5 min before each sample measurement was taken.

The percentage of fluorescence decrease was calculated using the following formula:

$$\% \text{ Fluorescence Decrease} = \frac{(F_{\text{Probe Only}} - F_{\text{Test Compound}})}{F_{\text{Probe Only}}} \times 100\%$$

Where F = fluorescence intensity.

10.6 Melting Temperature Experiment

The bacterial rRNA (3 μM) was first annealed in a pH 7.0 buffer before the SPOC probe (3 μM) was added. The experiment was carried out in a 8-microcell quartz cuvette with an optical path of 10 mm and a UV-2450 UV-Visible spectrophotometer from Shimadzu (Kyoto, Japan). The total sample volume was 120 μL and paraffin was added at the top to prevent solvent loss. Absorbance was monitored at 260 nm (data not shown) and 300 nm with a temperature ramp of 0.5°C / min to give the resultant absorption spectra. The melting points were determined using the peaks in the first derivative curves of the absorption spectra.

11. Supporting Information

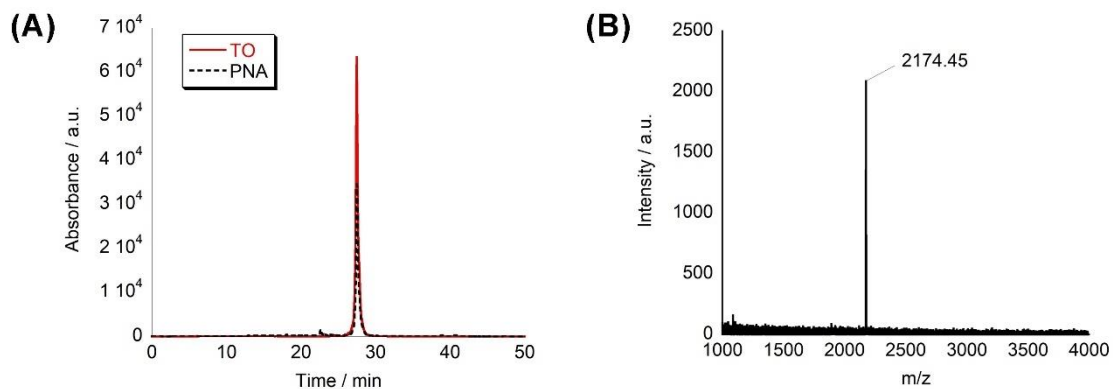


Figure S1. (A) HPLC chromatogram and (B) the corresponding MALDI-TOF-MS spectrum of purified SPOC probe with ATMND-C₂-NH₂.

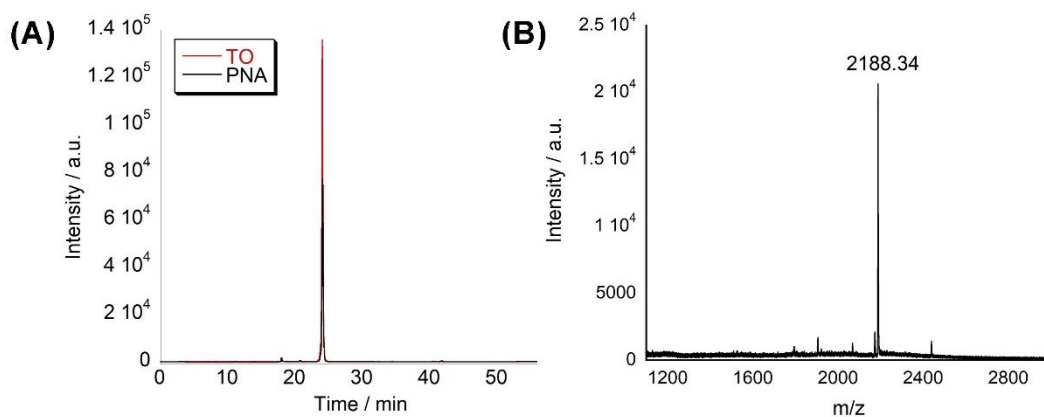


Figure S2. (A) HPLC chromatogram and (B) the corresponding MALDI-TOF-MS spectrum of purified SPOC probe with ATMND-C₃-NH₂.

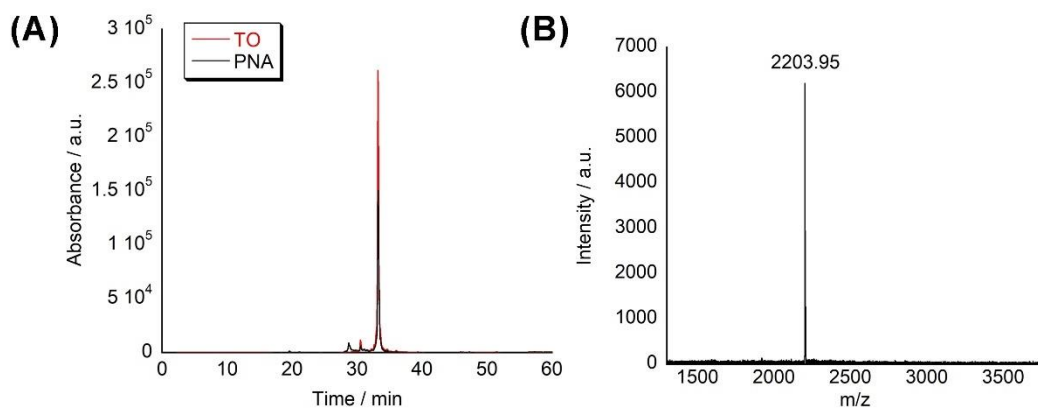


Figure S3. (A) HPLC chromatogram and (B) the corresponding MALDI-TOF-MS spectrum of purified SPOC probe with ATMND-C₄-NH₂.

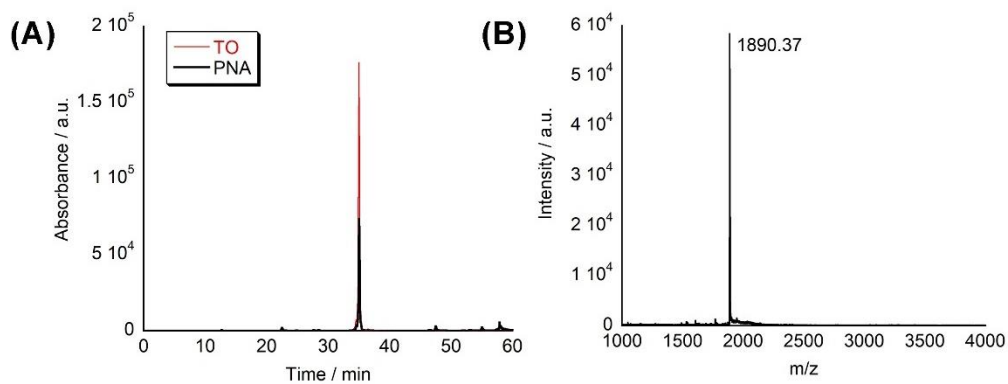


Figure S4. (A) HPLC chromatogram and (B) the corresponding MALDI-TOF-MS spectrum of the purified control probe.

Table S1 MALDI-TOF-MS data for the various probes.

Probe	Chemical Formula	Calculated MW [M] ⁺	Observed m/z
SPOC probe (ATMND-C ₂ -NH ₂)	C ₉₈ H ₁₂₆ N ₃₇ O ₂₀ S	2174.38	2174.45 [M] ⁺
SPOC probe (ATMND-C ₃ -NH ₂)	C ₉₉ H ₁₂₈ N ₃₇ O ₂₀ S	2188.41	2188.34 [M] ⁺
SPOC probe (ATMND-C ₄ -NH ₂)	C ₁₀₀ H ₁₃₀ N ₃₇ O ₂₀ S	2202.44	2203.95 [M+H] ⁺
Control Probe	C ₈₂ H ₁₀₆ N ₃₃ O ₁₉ S	1890.02	1890.37 [M] ⁺

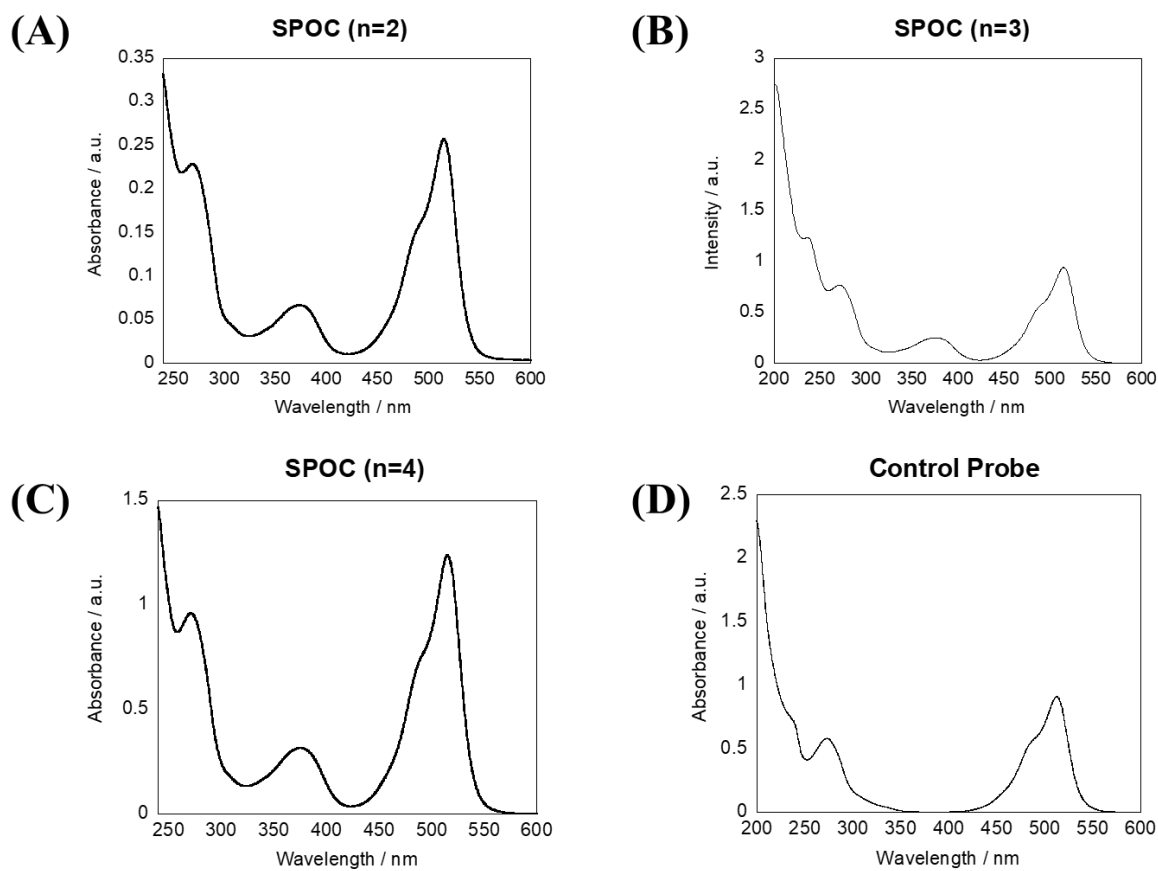


Figure S5. UV-Vis absorbance spectra of **(A)** SPOC ($n=2$), **(B)** SPOC ($n=3$), **(C)** SPOC ($n=4$) and **(D)** the control probe in MilliQ. Note that all the SPOC probes (A – C) contain the peak belonging to the ATMND moiety in the 330 – 400 nm range, while all the probes contain the TO peak (420 – 550 nm).

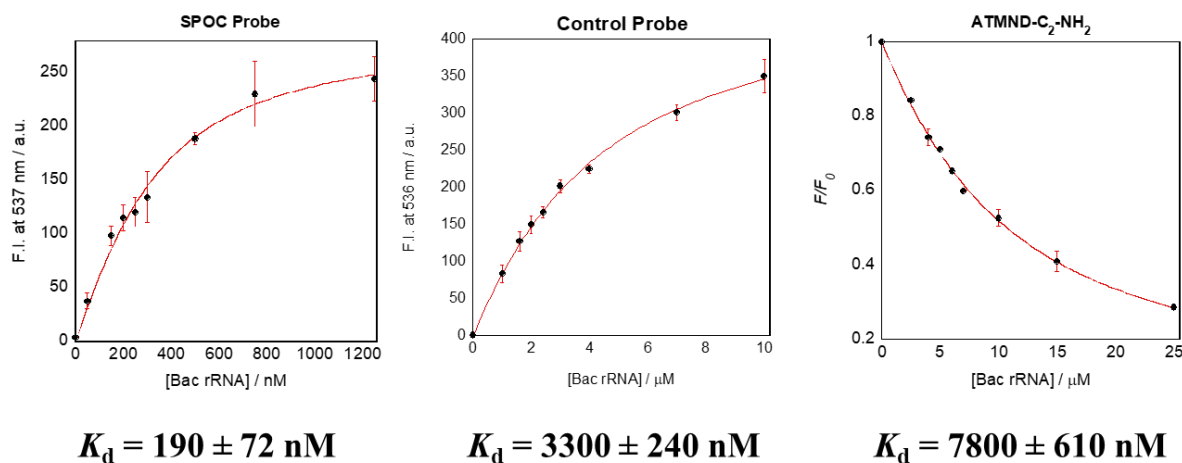


Figure S6. Concentration dependence curves of the SPOC probe, the control probe and ATMND-C₂-NH₂ with Bac rRNA at pH 7.0. The error bars were plotted from three independent experiments. [SPOC probe] = 250 nM, [Bac rRNA] = 0 – 1250 nM, Ex wavelength: 515 nm. [Control Probe] = 1.66 μM, [Bac rRNA] = 0 – 10 μM, Ex wavelength: 519 nm. [ATMND-C₂-NH₂] = 5 μM, [Bac rRNA] = 0 – 25 μM, Ex wavelength: 358 nm, F.I. values at 407 nm were used to plot the curve. Error bars were plotted from three independent measurements.

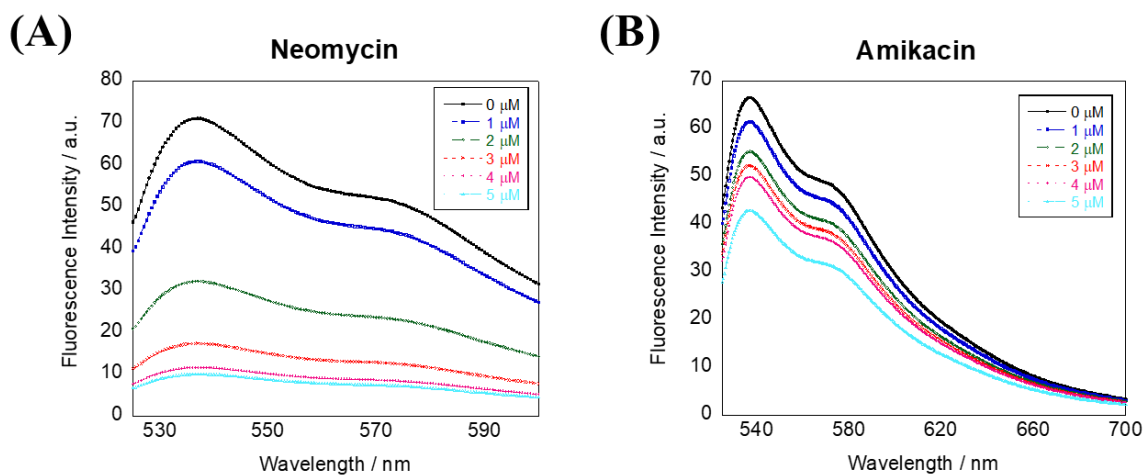


Figure S7. Fluorescence spectra of SPOC and Bac rRNA upon the addition of (A) neomycin and (B) amikacin. [SPOC] = [Bac rRNA] = 1 μM, [Neomycin or Amikacin] = 0 – 5 μM, Em band = 5 nm, Ex band = 5 nm, Sensitivity = Low, Ex wavelength = 513 nm, pH 7.0, 25°C.

12. References

- Barbieri, C., Kaul, M., & Pilch, D. (2007). Use of 2-aminopurine as a fluorescent tool for characterizing antibiotic recognition of the bacterial rRNA A-site. *Tetrahedron*, *63*(17), 3567-3574.
- Carter, A., Clemons, W., Brodersen, D., Morgan-Warren, R., Wimberly, B., & Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, *407*(6802), 340-348.
- Chen, Y., Fan, Y., Tien, C., Yueh, A., & Chang, R. (2018). The conserved stem-loop II structure at the 3' untranslated region of Japanese encephalitis virus genome is required for the formation of subgenomic flaviviral RNA. *PLOS ONE*, *13*(7), e0201250.
- Chittapragada, M., Roberts, S., & Ham, Y. (2009). Aminoglycosides: Molecular Insights on the Recognition of RNA and Aminoglycoside Mimics. *Perspectives In Medicinal Chemistry*, *3*, PMC.S2381.
- D'Souza, V., Dey, A., Habib, D., & Summers, M. (2004). NMR Structure of the 101-nucleotide Core Encapsidation Signal of the Moloney Murine Leukemia Virus. *Journal Of Molecular Biology*, *337*(2), 427-442.
- Dudek, M., Romanowska, J., Wituła, T., & Trylska, J. (2014). Interactions of amikacin with the RNA model of the ribosomal A-site: Computational, spectroscopic and calorimetric studies. *Biochimie*, *102*, 188-202.
- Forge, A., & Schacht, J. (2000). Aminoglycoside Antibiotics. *Audiology and Neurotology*, *5*(1), 3-22
- Garneau-Tsodikova, S., & Labby, K. (2016). Mechanisms of resistance to aminoglycoside antibiotics: overview and perspectives. *Medchemcomm*, *7*(1), 11-27.
- Gupta, P., Zengeya, T., & Rozners, E. (2011). Triple helical recognition of pyrimidine inversions in polypurine tracts of RNA by nucleobase-modified PNA. *Chemical Communications*, *47*(39), 11125.
- Han, T. (2022). Synthesis and evaluation of new triplex-forming PNA-based fluorogenic probes targeting the bacterial rRNA A-site. 修士論文.
- Hong, S., Harris, K., Fanning, K., Sarachan, K., Frohlich, K., & Agris, P. (2015). Evidence That Antibiotics Bind to Human Mitochondrial Ribosomal RNA Has Implications for Aminoglycoside Toxicity. *Journal of Biological Chemistry*, *290*(31), 19273-19286.
- Jana, S., & Deb, J. (2006). Molecular understanding of aminoglycoside action and resistance. *Applied Microbiology and Biotechnology*, *70*(2), 140-150.
- Kaul, M., & Pilch, D. (2002). Thermodynamics of Aminoglycoside-rRNA Recognition: The Binding of Neomycin-Class Aminoglycosides to the A Site of 16S rRNA. *Biochemistry*, *41*(24), 7695-7706.

- Köhler, O., Jarikote, D., & Seitz, O. (2004). Forced Intercalation Probes (FIT Probes): Thiazole Orange as a Fluorescent Base in Peptide Nucleic Acids for Homogeneous Single-Nucleotide-Polymorphism Detection. *Chembiochem*, 6(1), 69-77.
- Krause, K., Serio, A., Kane, T., & Connolly, L. (2016). Aminoglycosides: An Overview. *Cold Spring Harbor Perspectives In Medicine*, 6(6), a027029.
- Lee, M., Bottini, A., Kim, M., Bardaro, M., Zhang, Z., & Pellicchia, M. et al. (2014). Correction: A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication. *Chem. Commun.*, 50(83), 12578-12578.
- Lin, J., Zhou, D., Steitz, T., Polikanov, Y., & Gagnon, M. (2018). Ribosome-Targeting Antibiotics: Modes of Action, Mechanisms of Resistance, and Implications for Drug Design. *Annual Review of Biochemistry*, 87(1), 451-478.
- Macmaster, R., Zelinskaya, N., Savic, M., Rankin, C., & Conn, G. (2010). Structural insights into the function of aminoglycoside-resistance A1408 16S rRNA methyltransferases from antibiotic-producing and human pathogenic bacteria. *Nucleic Acids Research*, 38(21), 7791-7799.
- Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., & Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nature Structural & Molecular Biology*, 19(6), 560-567.
- Mergny, J., Lacroix, L., Han, X., Leroy, J., & Helene, C. (1995). Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an i-DNA Motif. *Journal of The American Chemical Society*, 117(35), 8887-8898.
- Mingeot-Leclercq, M., Glupczynski, Y., & Tulkens, P. (1999). Aminoglycosides: Activity and Resistance. *Antimicrobial Agents and Chemotherapy*, 43(4), 727-737.
- Neumann, K., Farnung, J., Baldauf, S., & Bode, J. (2020). Prevention of aspartimide formation during peptide synthesis using cyanosulfurylides as carboxylic acid-protecting groups. *Nature Communications*, 11(1).
- Ogle, J., Brodersen, D., Clemons, W., Tarry, M., Carter, A., & Ramakrishnan, V. (2001). Recognition of Cognate Transfer RNA by the 30 S Ribosomal Subunit. *Science*, 292(5518), 897-902.
- Pack, G., Wong, L., & Lamm, G. (1998). PK_a of cytosine on the third strand of triplex DNA: Preliminary Poisson-Boltzmann calculations. *International Journal Of Quantum Chemistry*, 70(6), 1177-1184.
- Pape T, Wintermeyer W, Rodnina MV. (2000) Conformational switch in the decoding region of 16S rRNA during aminoacyl-tRNA selection on the ribosome. *Nat Struct Biol.*, 7(2), 104-7.
- Ramakrishnan, V. (2002). Ribosome Structure and the Mechanism of Translation. *Cell*, 108(4), 557-572.

Ray, A., & Nordén, B. (2000). Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future. *The FASEB Journal*, *14*(9), 1041-1060.

Sato, T., Sato, Y., & Nishizawa, S. (2016). Triplex-Forming Peptide Nucleic Acid Probe Having Thiazole Orange as a Base Surrogate for Fluorescence Sensing of Double-stranded RNA. *Journal of The American Chemical Society*, *138*(30), 9397-9400.

Sato, Y., Rokugawa, M., Ito, S., Yajima, S., Sugawara, H., Teramae, N., & Nishizawa, S. (2018). Fluorescent Trimethylated Naphthyridine Derivative with an Aminoalkyl Side Chain as the Tightest Non-aminoglycoside Ligand for the Bacterial A-site RNA. *Chemistry - A European Journal*, *24*(52), 13862-13870.

Sato, Y., Miura, H., Tanabe, T., Okeke, C., Kikuchi, A., & Nishizawa, S. (2022). Fluorescence Sensing of the Panhandle Structure of the Influenza A Virus RNA Promoter by Thiazole Orange Base Surrogate-Carrying Peptide Nucleic Acid Conjugated with Small Molecule. *Analytical Chemistry*, *94*(22), 7814-7822.

Wang, C., Sato, Y., Kudo, M., Nishizawa, S., & Teramae, N. (2012). Ratiometric Fluorescent Signaling of Small Molecule, Environmentally Sensitive Dye Conjugates for Detecting Single-Base Mutations in DNA. *Chemistry - A European Journal*, *18*(31), 9481-9484.

Chapter 3

Light-Up Peptide Indicator for HIV-1 TAR RNA

1. Introduction to RNPs

When an RNA-binding protein (RBP) hybridizes with a RNA, it forms a ribonucleoprotein (RNP) complex. As simple as this interaction may seem, RNPs have been gaining attention as scientists discover more and more about how RNPs regulate gene expression and localisation in the cell. In fact, a census taken in 2014 counted 1,542 RBPs in the human genome that bind with all known classes of RNA (Gerstberger, Hafner & Tuschl, 2014). Recent developments of large-scale quantitative methods to discover new RNPs such as the use of systematic evolution of ligands by exponential enrichment (SELEX) to predict the RNA recognition elements (RRE) of RBPs (Galarneau & Richard, 2005) and deep sequencing techniques such as CLIP-seq (crosslinking and immunoprecipitation followed by sequencing) or complexomic approaches (Grad-seq) (Smirnov et al., 2016) have facilitated the discovery of many new RNPs.

The increase in the number of known RNPs has resulted in the discovery of more RNP bacterial and viral RNPs with therapeutic significance. For example, using the Grad-seq approach, Smirnov and his team were able to deduce that in *Salmonella*, the RNA chaperone ProQ post-transcriptionally regulates ~16% of its genome, making ProQ-small regulatory RNAs (sRNAs) a plausible drug target (Smirnov, Schneider, Hör & Vogel, 2017). In 2016, the crystal structure of the core domain structure of the nucleocapsid protein (NP) of the hantavirus encoded by Sin Nombre virus (SNV) and Andes virus (ANDV) was resolved, where a hydrophobic RNA binding crevice was discovered. Ligands that can bind to that crevice and thus inhibit RNP formation could potentially limit viral proliferation (Guo et al., 2016). AU-rich element-binding proteins (AUBPs) that bind to AU-rich elements (AREs) in the 3'-UTR of mRNAs have shown to be involved in cancer progression and the AUBP-ARE interaction could also be of therapeutic significance (Dolicka, Sobolewski, Correia de Sousa, Gjorgjieva & Foti, 2020). Human antigen R (HuR), an AUBP, was shown to have its interaction with AREs inhibited by small molecules CMLD-1 and CMLD-2 from high throughput screening of 6000 compounds, (Wu et al., 2015). Thus, RNPs started to be viewed as possible drug targets.

The potential of RNPs as druggable targets represents a significant expansion to the current druggable space as the search begins for small molecules that can interfere with dynamic protein-RNA surface (D'Agostino et al., 2019). To aid in the large-scale screening required, FID assays serve as an attractive high-throughput method. Furthermore, one could take advantage of the fact that the RBP in the RNP of interest is already a strong dsRNA binder, and thus design a truncated fluorescent peptide using the amino acid sequence of the wild-type RBP itself. This truncated peptide indicator idea is what we focused in on when designing LUPI.

2. Previous Peptide Indicators

This is by no means a new idea. In fact, in 2000, Matsumoto and team designed a fluorescent peptide indicator for the HIV-1 Transactivation response region (TAR) RNA by using the amino acid sequence of the arginine-rich motif (ARM) of the HIV-1 Tat peptide (Matsumoto, Hamasaki, Mihara & Ueno, 2000). The Tat protein binds to the TAR RNA to form an RNP that is essential in facilitating viral replication. Without the Tat-TAR interaction that stabilises the downstream replication, RNA polymerases stall near the viral promoter and the virus is unable to replicate (Aboul-ela, Karn & Varani, 1995). Therefore, a molecule that can inhibit the Tat-TAR interaction would likely work as an antiviral agent. Matsumoto utilised a fluorescence resonance electron transfer (FRET) system between terminally-linked fluorescein and rhodamine dyes in his Tat peptide probe, which he aptly named, FtatRhd. However, upon binding with TAR RNA, the fluorescence intensity of the probe only increased ~3 times. Furthermore, the poor photostability of the probe owing to the rhodamine made it challenging to work with. Despite its drawbacks, laboratories around the world have still been using this FtatRhd probe as recently as 2019 (Patwardhan, Cai, Newson & Hargrove, 2019). We thought that it was time that this proverbial dinosaur received a significant improvement.

Another existing peptide indicator was based on the boxB RNA- λ N protein complex that Jeong and his team designed (Jeong et al., 2013). In bacteriophage λ , interactions between the N-protein and the boxB stem-loop RNA element are essential in transcription anti-termination (Zhang, Lee, Zhao, Xia & Qin, 2010). The peptide indicator sequence used the first 22 amino acids of the λ N protein, which is also an arginine-rich motif (ARM). They aimed to create a peptide that had a lower binding affinity to the RNA as compared to the wild-type protein by substituting one amino acid with a pyrene fluorophore. The lower binding affinity would thus allow the indicator to be displaced by the wild-type protein. Upon hybridisation with the boxB RNA, the pyrene would respond fluorescently through photon-induced electron transfer (PET), as the distance between the pyrene moiety and the tryptophan residue increased.

3. Probe Design and Research Objectives

3.1 Light-Up Peptide Indicator (LUPI) Design

To put our own creative spin on things, we took inspiration from the previously mentioned tFIT probe design that our lab had developed (Sato, Sato & Nishizawa, 2016). In the tFIT probe design, thiazole orange (TO) was used as a surrogate nucleobase, capable of base-pairing with any opposing RNA nucleobase in the dsRNA. We wondered if we could similarly use TO as a *surrogate amino acid* in our peptide indicator. To the best of our knowledge, TO has never been reported to be used in this way before. In this probe design, TO would have to intercalate into the dsRNA target while keeping the peptide's function unchanged, thus serving both as a fluorophore and a surrogate. We decided to call this probe design the light-up peptide indicator (LUPI).

In order to prove if this concept worked, we decided to use the **HIV-1 Tat protein – TAR RNA model** as it is an incredibly well-known and well-used model. We followed Matsumoto's lead and used the arginine-rich motif in the Tat protein as the base amino acid sequence for our Tat peptide as this 9-amino-acid sequence was shown to be responsible RNA binding (Roy, Delling, Chen, Rosen & Sonenberg, 1990).

3.2 Research Objectives

With the LUPI probe design, we aimed to develop a probe that had a large fluorescence response, that had good binding affinity with TAR RNA, and that could be used as an FID indicator to search for Tat-TAR inhibitors.

3.3 Choosing which Amino Acid to Substitute with TO

There were two requirements that had to be met when choosing which amino acid to substitute with TO – the amino acid chosen could not be essential for TAR binding, and the amino acid chosen had to be close in proximity to TAR RNA upon binding such that TO would be able to intercalate into the RNA.

According to previous reports, Tat protein binding with TAR RNA is mediated by a single arginine, either at position 52 or 53 (Calnan, Tidor, Biancalana, Hudson & Frankel, 1991). As such, neither Arg52 or Arg 53 should be substituted with TO. Moreover, using EPR spectroscopy, Edwards and team found that C-terminal arginines (53, 55, 56 and 57) are needed for rigid Tat-TAR complex formation and should also not be replaced with TO (Edwards, Robinson & Sigurdsson, 2005).

Gln54 was shown to be uninvolved in TAR RNA binding and could be deleted without adverse effects (Delling et al., 1991). NMR-resolved structures of HIV-1 Tat protein – TAR RNA complex is difficult due to a cysteine rich region in the Tat activation domain. Instead, scientists have looked at other analogous Tat proteins from related viruses to gain insights to HIV Tat-TAR (Greenbaum, 1996). By looking at a complex of TAR RNA and a related Tat protein from the EIAV (equine infectious anemia virus), the “country cousin” of the HIV-1 variant which is related to HIV by its structure (Leroux, Cadore & Montelaro, 2004) but lacks the cysteine-rich domain, it seemed that Gln54 was located in close proximity to the RNA (Anand, Schulte, Vogel-Bachmayr, Scheffzek & Geyer, 2008).

As such, we decided to substitute Gln54 with TO. We were concerned about whether substituting a neutrally-charged Gln with a positively-charged TO moiety would negatively affect the TAR binding, however, previous reports mention that a 9-mer homo-arginine peptide, where each arginine carried one positive charge, was still able to bind specifically to TAR RNA (Calnan, Tidor, Biancalana, Hudson & Frankel, 1991), which meant that adding an additional positive charge to the Tat peptide would have no adverse effects on the selectivity of the peptide.

3.4 LUPI Amino Acid Sequence

The final amino acid sequence for LUPI was N' – **R K K R R Dap(TO) R R R** – C' (9-mer).

A Dap amino residue was chosen as the linker for TO-C₁-COOH because an amino acid with a -NH₂ functional group was needed for TO coupling. Amongst the various amino acids with -NH₂ functionality (Dap, Dab and Lys), Dap was chosen as it is the shortest. We reasoned that for TO to act as a surrogate amino acid, the final length of the Dap-TO residue should be similar to that of twenty essential amino acids, of which arginine is one of the longest. In order to mimic the length of arginine, any amino acid other than Dap would be too long.

3.5 Other Probes for Comparison

To better understand some of the structural aspects of LUPI, three other probes were synthesised for comparison – the conventional probe, the Lys-linker probe and the original FtatRhd probe.

The conventional probe had TO coupled to the N-terminal of the Tat peptide, similar to how conventionally, the fluorophore would be terminally-linked. Comparison with the conventional probe would help us to

understand the importance of the position of the fluorophore. TO was coupled to the probe via a Lys linker as this would also have been the conventional choice of linker.

The Lys-linker probe contains TO coupled in the same position as the LUPI probe, but with a Lys linker instead of a Dap linker. Comparison with the Lys-linker probe would help us to understand the importance of the linker length of the fluorophore.

The FtatRhd probe was ordered from HiPep Laboratories, Kyoto, so that we could directly compare LUPI with the original Tat peptide indicator (Matsumoto, Hamasaki, Mihara & Ueno, 2000). Three alanine residues were added to each terminal in order to maximise the FRET distance between the two fluorophores upon hybridising with the dsRNA.

Table 1. Table of the amino acid sequences of all the probes used in this research.

Probe	Probe Sequence
LUPI	R K K R R Dap(TO) R R R
Conventional Probe	Lys(TO) R K K R R Q R R R
Lys-Linker Probe	R K K R R Lys(TO) R R R
FtatRhd	α -FAM-A A A R K K R R Q R R R A A A K(ϵ TAMRA)-NH ₂

3.6 Comparing two types of Thiazole Orange

Thiazole orange is made up of a quinoline moiety attached to a benzothiazole through a monomethine link. As such the carboxyl group of TO-C₁-COOH can be attached either to the quinoline (TO_Q) or to the benzothiazole (TO_Z).

As we did not know which version of TO would allow for better intercalation into the dsRNA, we decided to synthesise two LUPI probes, one with TO_Q and one with TO_Z, to compare the fluorescence intensity.

TO_Q is the usual TO used for our tFIT probes and was available in-house. TO_Z was synthesised as previously reported (Carreon, Mahon, & Kelley, 2004). The overall synthesis scheme is shown in Fig. 2.

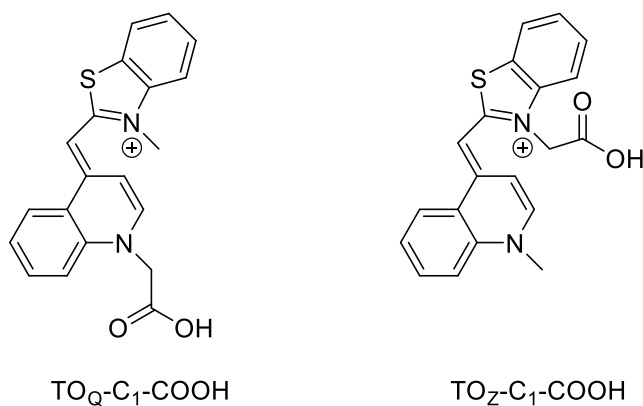


Figure 1. Chemical structures of TO_Q-C₁-COOH and TO_Z-C₁-COOH.

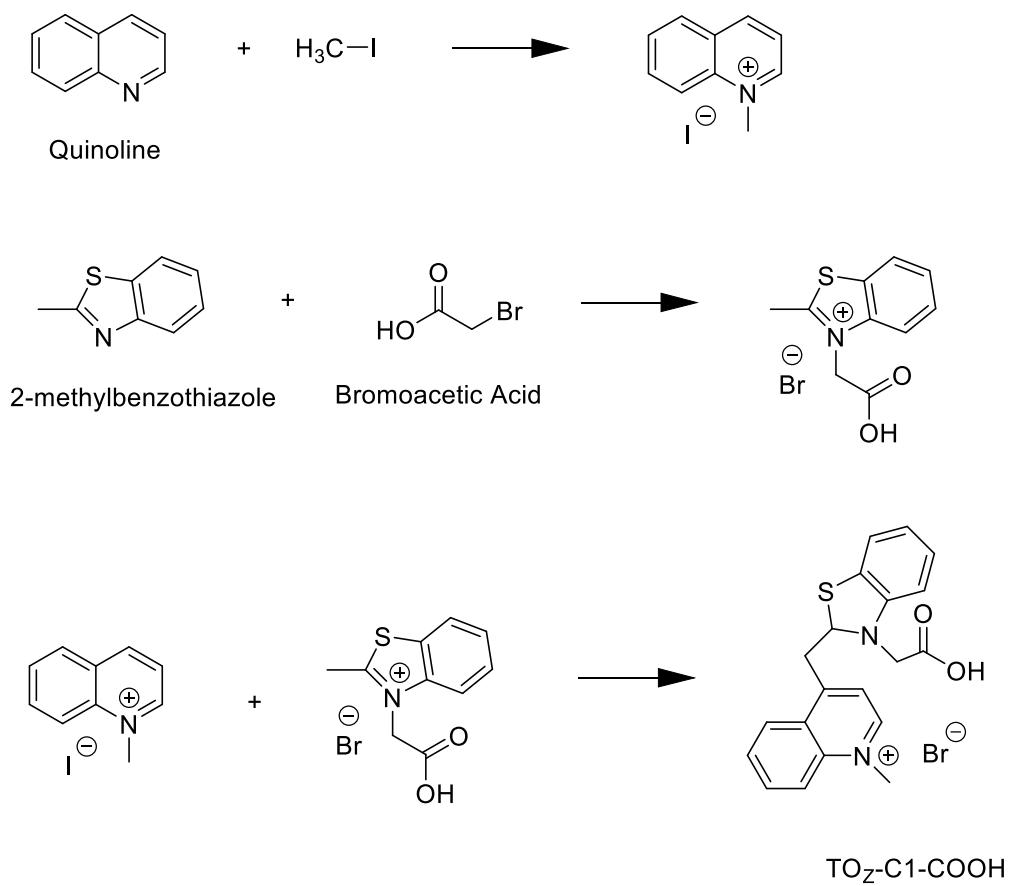


Figure 2. Overall scheme for the synthesis of $\text{TO}_z\text{-C}_1\text{-COOH}$

4. Significance of the HIV-1 Tat protein – TAR RNA RNP Complex

The human immunodeficiency virus type 1 (HIV-1) is a virus that was first recognised to cause the Acquired Immune Deficiency Syndrome (AIDS) 39 years ago (Barré-Sinoussi et al., 1983). Since then, it has taken the life of more than 20 million people (Nathans et al., 2008). Unsurprisingly, a lot of research has gone into elucidating the mechanisms of pathogenesis and viral replication.

The Tat (Trans Activating Factor) protein is a regulatory protein of HIV that is essential for viral transcription and replication. Even though the HIV virus is known for its high levels of mutation, which is one of the main challenges of developing a HIV vaccine, the Tat protein is relatively well-conserved (Debaisieux, Rayne, Yezid & Beaumelle, 2011). The Tat protein consists of five different domains: the N-terminal domain, the cysteine-rich domain, the core, the basic domain (that consists of the arginine rich-motif, ARM) and the C-terminal domain, and it is the basic region is responsible for the binding with TAR-RNA (Pugliese, Vidotto, Beltramo, Petrini & Torre, 2005).

During the start of viral transcription, RNA Polymerase II (Pol II) stalls after transcribing ~50 nucleotides which contains the TAR region. The Tat protein recruits the host super elongation complex (SEC), which consists of positive elongation factor b (P-TEFb), composed of CDK9 and Cyclin T1 (CycT1), the transcriptional elongation factors ELL2 and ENL/AF9, and the ~1,200-amino acid scaffold proteins AFF1 and/or AFF4. This Tat-SEC complex binds to TAR RNA to release the stalled Pol II which restarts the transcription. Hence, the Tat-TAR interaction is essential to produce full-length viral transcripts that are needed for proper viral gene expression and replication (Peterlin & Price, 2006; He et al., 2010; Schulze-Gahmen et al., 2016)

4.1 Binding of Tat protein to TAR RNA

Tat binds to TAR RNA through the UCU bulge. This was deduced as deletions of the uridines in the bulge led to a drop in binding affinity with Tat (Dingwall et al., 1990), and disruption of nearby base-pairing to affect the bulge structure abolished Tat binding (Karn, 1999). NMR studies of TAR RNA bound to a Tat peptide revealed how both TAR RNA and the Tat protein / peptide undergo conformational changes to bind together specifically. Free, unbound TAR RNA has a flexible structure. Upon binding, one arginine residue from the Tat ARM displaces U23 out of the helix, resulting in a binding pocket for that arginine side chain in the major groove, together with the G26•C39 base-pair, widening the major groove that facilitates more hydrogen bonds between G26 and the arginine (Aboul-ela, Karn & Varani, 1995; Brodsky & Williamson, 1997; Karn, 1999).

It has been proposed that there is base-triple formed between U23•A27•U38 (Puglisi, Tan, Calnan, Frankel & Williamson, 1992) to facilitate opening up the major groove even more, however, the available data does not seem to unanimously agree on that (Aboul-ela *et al.*, 1995).

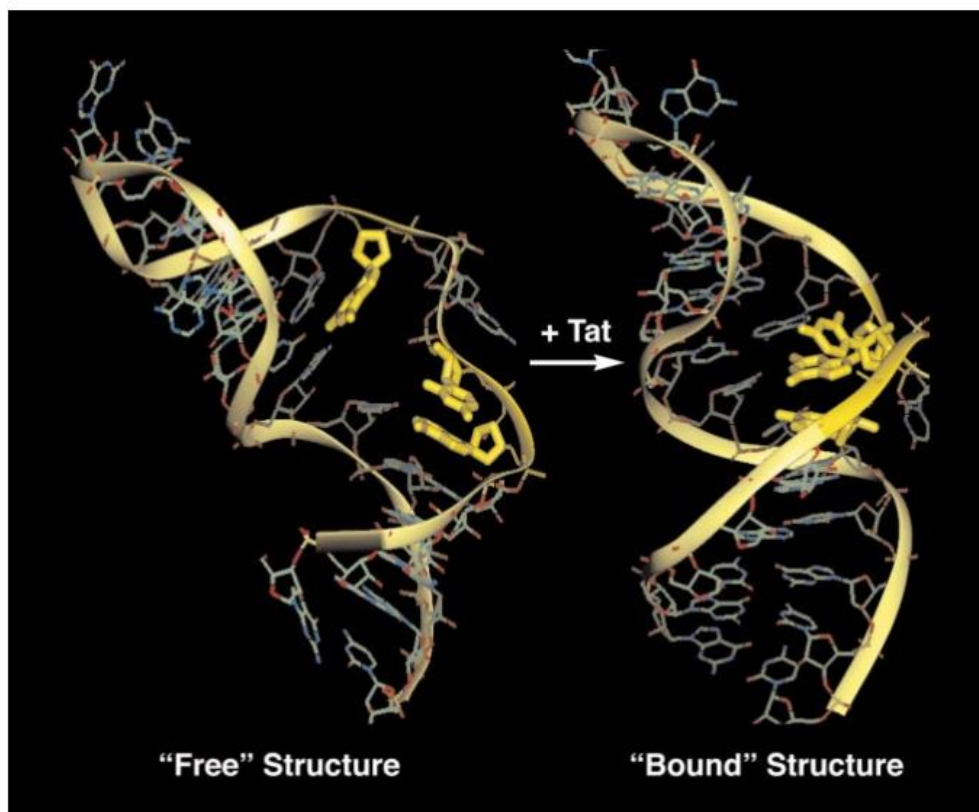


Figure 3. Major groove view of free TAR RNA (left) and TAR RNA in its Tat-bound configuration (right) (Karn, 1999). Residues A22, U23 and G26 are bolded. Upon binding, TAR RNA changes its conformation to resemble an A-form helix, and U23 is displaced out of the helix. Reprinted from *Journal of Molecular Biology*, 293, Jonathan Karn, Tackling tat, 235 - 254., Copyright (1999), with permission from Elsevier.

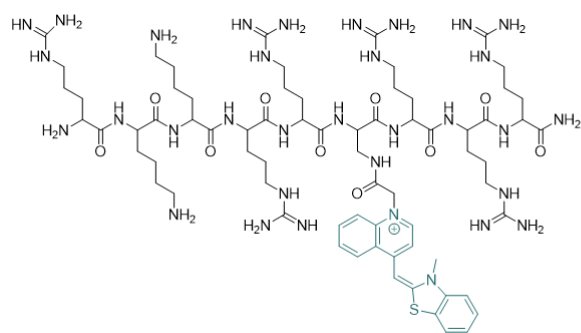
5. Probe Synthesis

LUPI, the conventional probe and the Lys-linker probe were made using the Biotage[®] Initiator+ Alstra[™] automatic microwave peptide synthesiser (Biotage, Uppsala, Sweden) using solid phase peptide synthesis with Fmoc chemistry. Peptide residues were purchased from Watanabe Chemical Industries Ltd. (Hiroshima, Japan).

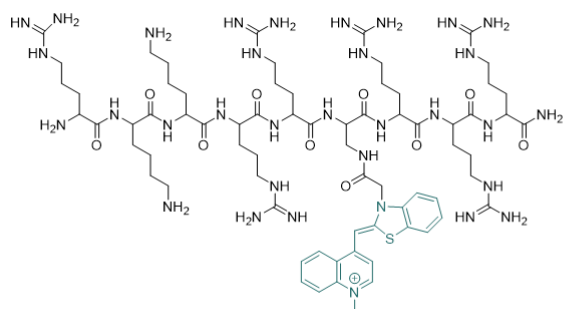
The arginine residue chosen was protected by a 2,2,4,6,7-pentamethylidihydrobenzofuran-5-sulfonyl group (Pbf) as it would offer more efficient peptide synthesis (Carpino et al., 1993). However, the Pbf protecting group is quite bulky, and thus might sterically hinder the peptide coupling of arginine. As such, for all the arginine residues, doubling coupling was done to ensure that arginine would be successfully coupled.

After the peptide was synthesised, thiazole orange (both TO_Q-C₁-COOH and TO_Z-C₁-COOH) would be manually coupled as per the protocol stated in Chapter 2.

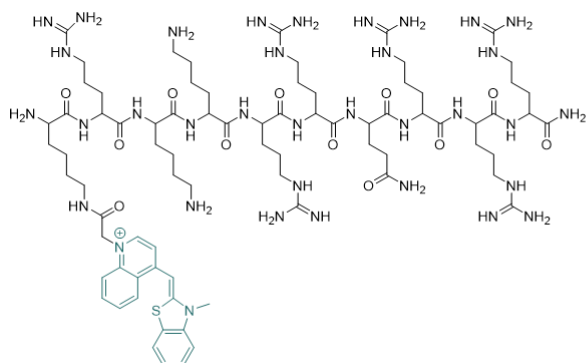
The synthesised probe was then confirmed by MALDI-TOF mass spectrometry and purified using reverse-phase HPLC. (Mass spectrum and HPLC chromatograms are available in the supporting information)



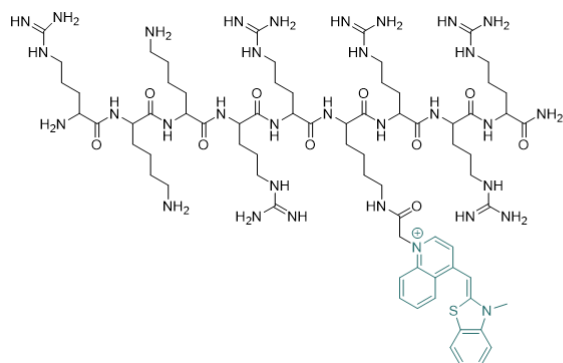
LUPI (TO_Q)



LUPI (TO_Z)



Conventional Probe



Lys-Linker Probe

Figure 4. Chemical structure of all the synthesised TO probes. The TO moiety has been highlighted in teal.

6. Fluorescence Light-Up Experiments

6.1 Comparing the two TOs

The first thing to be optimised was which version of TO should be used. For LUPI to serve as a good FID indicator, the change in fluorescence intensity upon hybridisation with the dsRNA should be maximised. As seen from Fig. 5, the LUPI carrying TO_Q had a fluorescence intensity ~3-4 times higher than the probe carrying TO_Z.

However, more interestingly, the different versions of TO resulted in different degree of selectivity between the probes. As seen in Fig. 6, LUPI carrying TO_Q was much more selective to TAR RNA compared to Bacteria A-site rRNA (Bac rRNA) and Influenza Promoter region viral RNA (Influenza vRNA). This is mostly due to the fact that LUPI carrying TO_Q resulted in much higher fluorescence intensity with TAR RNA as compared to TO_Z, as the response of both probes with Bac rRNA and Influenza vRNA was approximately the same, indicating that TO_Q and TO_Z both intercalate to the same degree in the non-target dsRNAs.

In the case of nucleic acids, TO_Q is preferred over TO_Z as TO_Q can bury deeper into the dsRNA, resulting in a higher change in fluorescence intensity (Carreon, Mahon, & Kelley, 2004). However, TO_Q intercalated into all three RNAs in the same way, then we should see TO_Q having higher F.I. with all three dsRNAs compared to TO_Z. The fact that the response of both TO_Q and TO_Z to the non-target dsRNAs are equivalent indicates that peptide-coupled TO_Q generally intercalating deeper into dsRNA does not seem to hold true in this case. Rather, the similarity in fluorescence response seems to imply that the binding of LUPI to the non-targets is non-specific and is driven by electrostatic interaction that results in TO_Q and TO_Z intercalating to the same degree.

Unlike non-specific hybridisation with the non-targets, it is well known that the binding of the Tat peptide with TAR RNA is very specific, inducing a conformational change in TAR RNA in the UCU bulge region (Aboul-ela, 1996). It can thus be inferred that this conformational change is causing the huge difference in fluorescence intensities between the two TO versions, and that it favours TO_Q intercalation over TO_Z intercalation.

Since TO_Q resulted in the higher fluorescence response, TO_Q was chosen as the version for which the conventional and Lys-linker probes were coupled with.

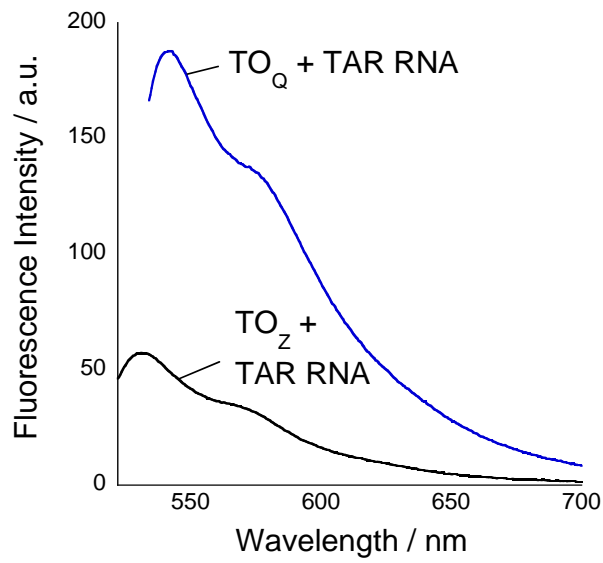


Figure 5. [LUPI] = [TAR RNA] = 50 nM. Excitation wavelength = 512 nm (TO_Z) and 524 nm (TO_Q). PBS Buffer, 25 °C.

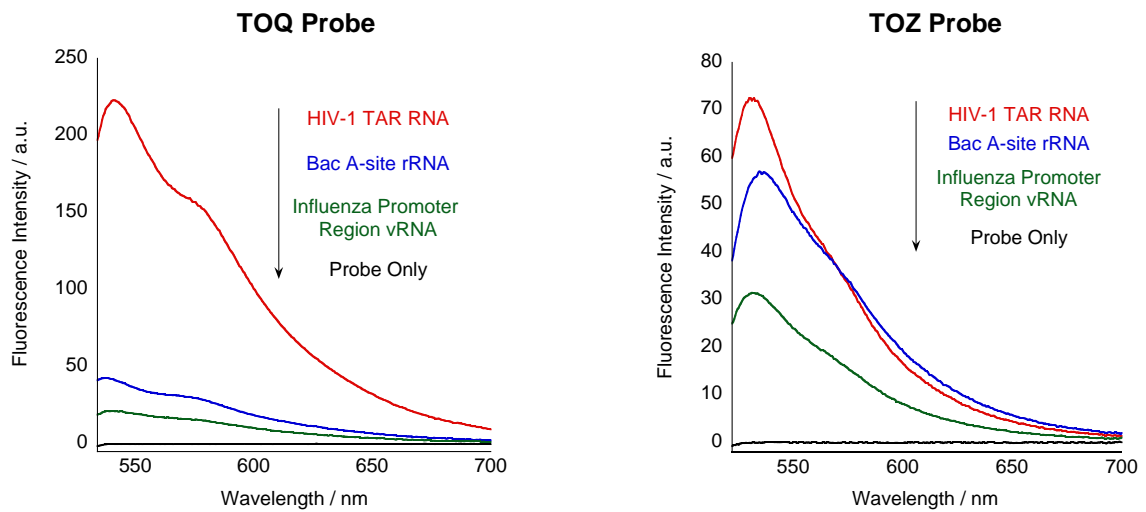


Figure 6. Comparing the selectivity between LUPI carrying TO_Q and LUPI carrying TO_Z. [Probe] = [RNA] = 50 nM. Excitation wavelength = 512 nm (TO_Z) and 524 nm (TO_Q). PBS Buffer, 25 °C.

6.2 Comparing the Fluorescence Intensity amongst All the Probes

For all the TO-carrying probes (LUPI, conventional and Lys-linker probes), in the absence of RNA, there was negligible fluorescence response. This is expected as un-intercalated TO is non-emissive and gives virtually zero background signal. In contrast, FtatRhd shows some fluorescence response even in the absence of RNA because of its FAM-TAMRA fluorophore system. FAM (fluorescein) is always emissive, and thus although it is slightly quenched by TAMRA when the two fluorophores are in close proximity, there will still be some background fluorescence signal, which is one of the drawbacks of a two-fluorophore system.

Upon the addition of TAR RNA, all the probes show an increase in fluorescence intensity, but LUPI shows the largest change in fluorescence, ~3-4 times higher than the other TO-carrying probes, and ~100-fold higher than FtatRhd.

This large change in fluorescence is impressive and tells us that not only is the version of TO important in maximising the fluorescence intensity, the position and linker length of TO all factor in as well.

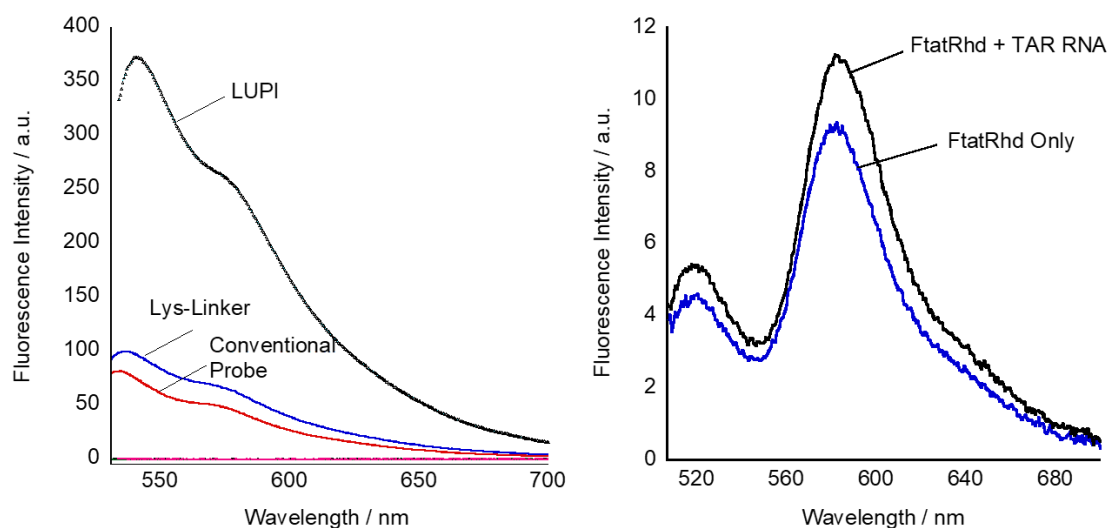


Figure 7. Fluorescence spectra of all the probes with TAR RNA. [Probe] = [TAR RNA] = 100 nM. Excitation wavelength = 497 nm (FtatRhd), 519 nm (Conventional and Lys-Linker Probe), 524 nm (LUPI), PBS Buffer, 25 °C.

6.3 Fluorescence Quantum Yield of LUPI

Next, we wanted to determine the fluorescence quantum yield of LUPI. The fluorescence quantum yield is a measure of the ratio of number of photons emitted by the substance to the number of photons absorbed by the substance. A value of 1 means that all the photons absorbed are emitted, while a value closer to 0 indicates that not all the photons absorbed are emitted.

$$\text{Fluorescence Quantum Yield, } \Phi_F = \frac{\text{Number of photons emitted}}{\text{Number of photons absorbed}}$$

There are two methods to determine the fluorescence quantum yield of a substance – a relative method, where the quantum yield is calculated relative to a standard substance, or an absolute method, where the actual number of photons are measured. I started out by wanting to determine the relative fluorescence quantum yield of LUPI, using fluorescein as the standard material, however, those measurements gave a value of 0.99, which was impossibly high, even higher than that of fluorescein (0.95) (Brannon & Magde, 1978).

Thus, we decided to measure the absolute fluorescence quantum yield with a quantum yield spectrometer and the determined values are in the table below.

Table 2. Table of values of the absolute fluorescence quantum yield of LUPI.

Absolute Fluorescence Quantum Yield (Φ_F)	
LUPI alone	0.00567
LUPI + TAR RNA	0.614

Upon the addition of TAR RNA, the fluorescence quantum yield increases by 108-fold, which accounts for the incredibly light-up response that LUPI exhibits. In the many probes that our laboratory has designed that has used TO as a fluorophore, LUPI showcases the highest Φ_F yet. Even the tFIT probe only has a Φ_F of 0.48, 20% lower compared to LUPI (Sato, Sato & Nishizawa, 2017), highlighting just how a good a light-up response LUPI has.

7. Circular Dichroism Experiments

To prove that LUPI's light-up response was indeed coming from the binding between LUPI and TAR RNA, circular dichroism experiments were conducted.

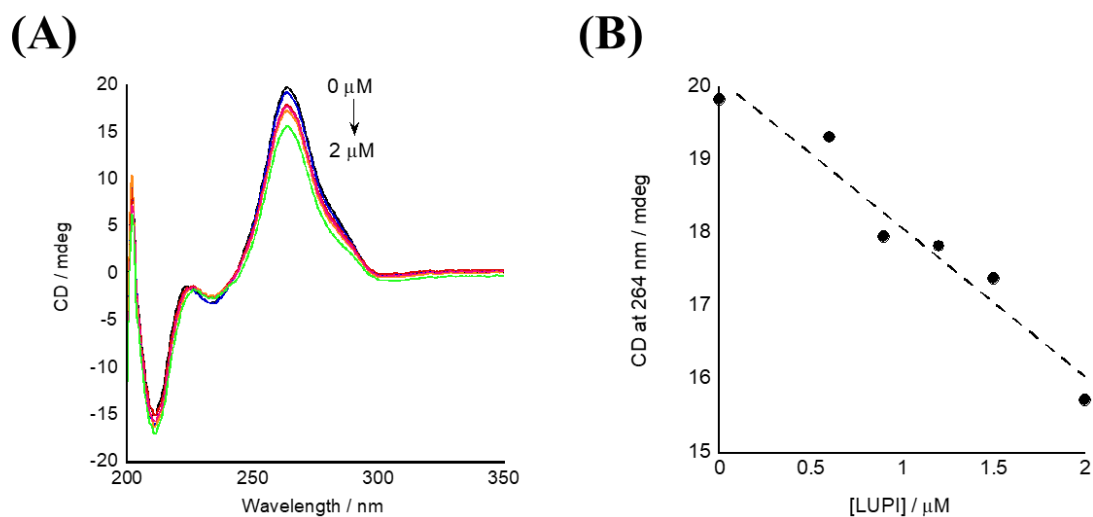


Figure 8. (A) CD spectra of TAR RNA (2 μM) with LUPI (0 – 2 μM), PBS Buffer, 25°C. (B) Plot of LUPI concentration against the CD response at 264 nm.

As seen in Fig. 8, without addition of LUPI, TAR RNA's CD spectrum matches that of a typical alpha-helical structure, and even with the addition of LUPI, the shape does not significantly change, implying that the LUPI-RNA complex is still in the A-form (Metzger et al., 1996). As more and more LUPI is added, the peaks at 264 nm and 210 nm decrease in intensity. According to previous reports, the peak around 260 nm comes from the RNA as the peptide does not absorb in that region (Tan & Frankel, 1992), while the peak at 210 nm most likely consists of contributions from both the RNA and the peptide. At 264 nm, the concentration-dependent conformational change (Fig. 8B) as more and more LUPI is added serves as solid evidence that LUPI is indeed binding to TAR RNA.

Moreover, as the CD spectra obtained matches that of a HIV-1 Tat peptide fragment that comprises the amino acids 32-72 of the Tat protein (Metzger et al., 1996), we can conclude that LUPI binds in a similar manner to previously reported Tat peptides.

8. Binding Affinity Experiments

The binding affinity of the various probes was also investigated by titration the probes against increasing concentrations of TAR RNA. The K_d value of LUPI with TAR RNA was determined to be 1.0 ± 0.6 nM, which can rival that of even the wild-type Tat protein (2 – 8 nM) (Chaloin et al., 2005). Although there are reports of cyclic Tat peptides that bind at even lower K_d values (Shortridge et al., 2018), this makes LUPI one of the strongest TAR RNA binders.

In contrast, FtatRhd was determined to have a K_d value of 360 ± 78 nM, which is two orders of magnitude higher than that of LUPI. The experimentally determined value of 360 nM is in good agreement with the reported value of 286 nM (Matsumoto, Hamasaki, Mihara & Ueno, 2000). We were unable to fit a curve through the data of the conventional probe and the Lys-linker probe, which indicates that these two probes have K_d values much higher than LUPI's. (Spectra shown in supporting information)

Binding affinity experiments were also done with the non-target Bac rRNA. However, in the case of Bac rRNA, at lower concentrations (< 1 eq. of LUPI's concentration), there was no perceivable pattern in the measured fluorescence intensity. It took concentrations twice as high as TAR RNA in order to see a concentration-dependent increase in fluorescence. This indicates that at lower concentrations, LUPI is binding non-specifically to Bac rRNA, which explains why there was no pattern to the binding. However, with excess Bac rRNA added (up to 10 eq.), concentration dependency could eventually be established, with the K_d value calculated to be 50 ± 15 nM. This supports the lack of selectivity of LUPI to Bac rRNA.

Table 3. Table of K_d values of probes with TAR RNA

Probe	K_d / nM
LUPI	1.0 ± 0.6
FtatRhd	360 ± 78
Conventional Probe	N.D.
Lys-Linker Probe	N.D.
Tat protein (Chaloin et al., 2005)	$\sim 2 - 8$

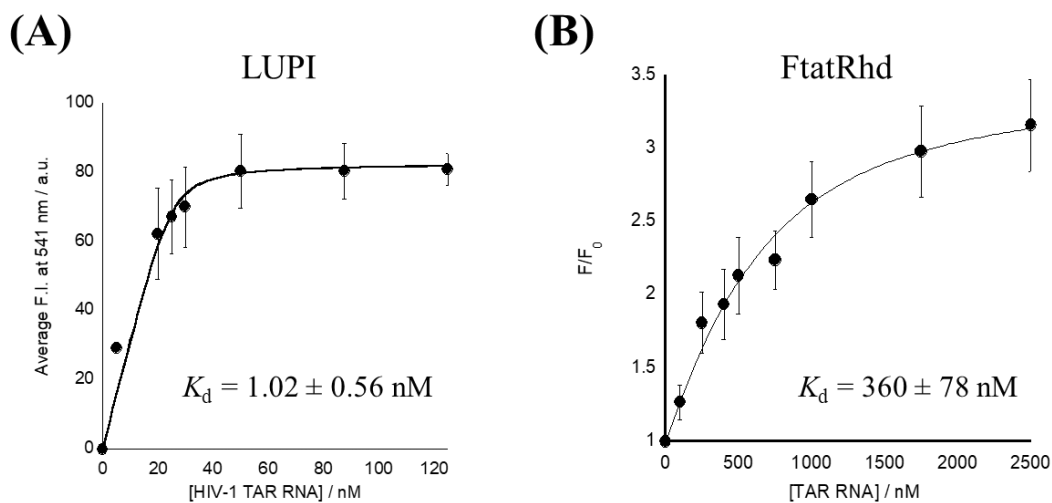


Figure 9. Concentration dependence curves of LUPI and FtatRhd with TAR RNA and the corresponding K_d values. Error bars were plotted from three separate experiments. **(A)** [LUPI] = 25 nM, [TAR RNA] = 0 – 125 nM, Excitation wavelength = 524 nm, PBS Buffer, 25°C. **(B)** [FtatRhd] = 500 nM, [TAR RNA] = 0 – 2500 nM, Excitation wavelength – 497 nm, PBS Buffer, 25°C.

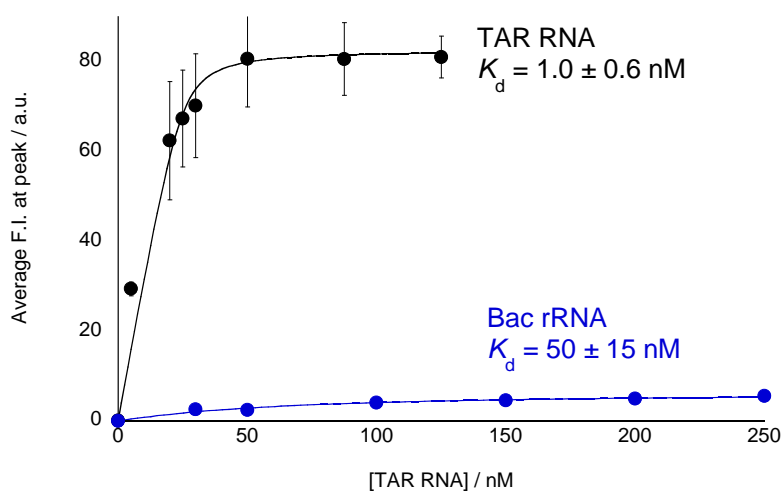


Figure 10. Concentration dependence curve of LUPI with TAR RNA and Bac rRNA. [LUPI] = 25 nM, [TAR RNA] = 0 – 125 nM, [Bac rRNA] = 0 – 250 nM. Excitation wavelength = 524 nm, PBS Buffer, 25°C.

The inability to fit a curve for the conventional and Lys-linker probe suggests that the K_d value of those two probes are at least two orders of magnitude higher than LUPI and shows that both the position of TO and the linker length of TO play important parts in modifying the K_d of LUPI.

It could be that these two probes are binding equally well to TAR RNA, but that TO is just not intercalating into the RNA. This is one of the drawbacks of LUPI, where the binding is monitored solely by TO's fluorescence. Either way, the position and linker length of TO are important parameters to optimise in order to develop a LUPI probe that works well in this system. As both probes carry TO in more flexible conformations (the terminal is more flexible than the middle of the peptide chain, and a longer linker length results in TO having more flexibility), it seems likely that a more rigid TO moiety allows for better intercalation.

8.1 Limit of Detection of LUPI and TAR RNA

As LUPI binds to TAR RNA at a low K_d value of 1 nM, we decided to measure the limit of detection (LOD) of LUPI and TAR RNA.

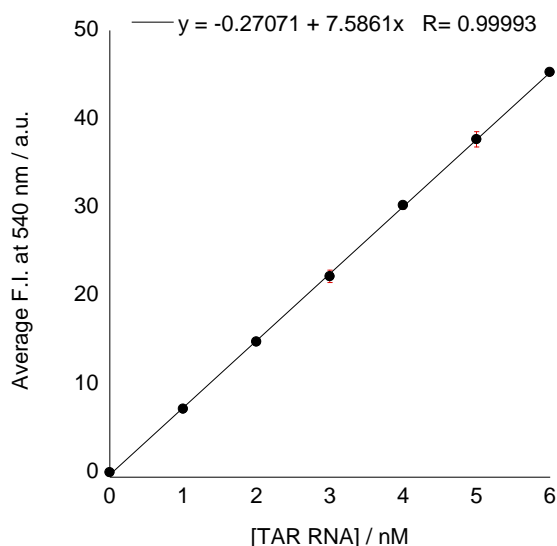


Figure 11. Calibration Curve of LUPI and TAR RNA. [LUPI] = 25 nM, [TAR RNA] = 0 – 6 nM, pH 7.4, 25°C. Error bars were plotted from three independent experiments.

$$\text{LOD} = 3(S_{\text{blank}} / \text{Slope}),$$

where S_{blank} is the standard deviation of the blank, and the slope refers to the slope of the calibration curve.

Using the above equation, the LOD of LUPI (25 nM) for TAR RNA was calculated to be **18 pM**, which is incredibly low. This highlights LUPI's potential to be used in other applications, and not just as an FID indicator.

9. Selectivity Experiments

As previously explained, when the Tat protein binds to TAR RNA, the basic ARM induces a conformational change in TAR RNA which places the functional groups of the RNA that is recognised by the protein (especially in G26 and U23) in a specific spatial arrangement (Aboul-ela, 1996; Brodsky & Williamson, 1997). This is the basis, but not the only factor, for the surprising but undeniable selectivity of the Tat peptide with TAR RNA.

As seen in Fig. 12A, when plotted against the fluorescence intensity of each probe, LUPI shows the highest fluorescence intensity towards TAR RNA. When the fluorescence responses are normalised against $F_{\text{TAR RNA}}$ (Fig. 12B), LUPI shows the best selectivity to TAR RNA over bacterial ribosomal RNA A-site (Bac rRNA) and the Influenza viral RNA promoter region (Influenza vRNA) out of all four probes.

It can be observed that although the Lys-linker probe shows selectivity to TAR RNA, its response to both Bac rRNA and Influenza vRNA is not significantly different. This contrasts with the rest of the probes, where their response to Bac rRNA is higher compared to Influenza vRNA. This lack of specificity of the Lys-linker probe is because of the longer linker length of TO, which leads to a greater degree of flexibility of the fluorophore. Although the position of TO in the probe is the same as that of LUPI, just changing the linker length dulls its ability to discriminate between two non-targets.

We were worried that the substitution of a neutrally-charged Q54 with a positively-charged thiazole orange could have adversely affected the selectivity of the Tat peptide. These results clearly show that the substitution did not abolish the Tat peptide's selectivity toward TAR RNA and that TO can successfully be used as a surrogate amino acid. These results are unsurprising, it has been reported that Tat-TAR recognition is mediated by a single arginine residue. Tao showed that a single arginine residue (in the context of a longer chain of Lys residues, 9-mer) was able to bind to TAR RNA with the same binding affinity and specificity as the wild-type Tat peptide (9-mer, ARM) (Tao & Frankel, 1992), indicating that as long as the single arginine was present, the other amino acids were substitutable. However, to the best of our knowledge, ours is the first report where TO has been used as an amino acid surrogate.

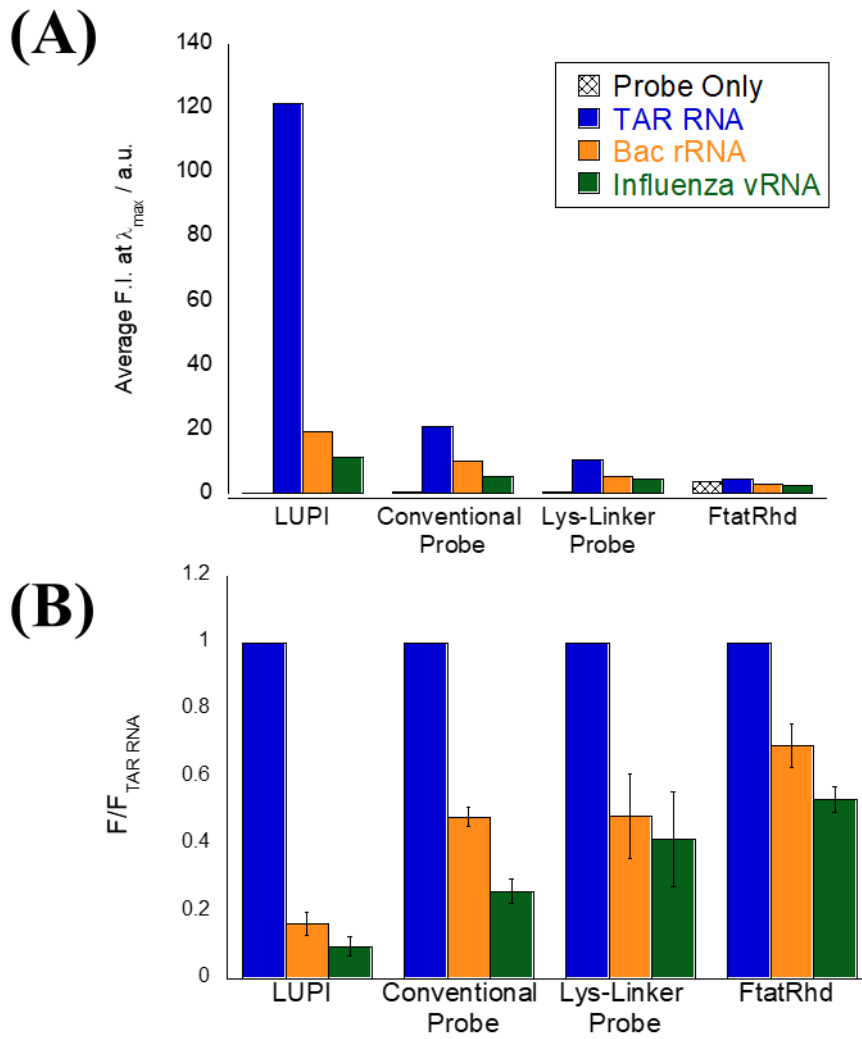


Figure 12. Comparing (A) fluorescence intensity or normalised fluorescence responses (B) between all four probes against three different RNAs. [Probe] = [RNA] = 50 nM.

9.1 Salt Effect Experiments

9.1.1 Salt Effect Experiments of LUPI with TAR RNA

While it is common knowledge in the science community for Tat to bind specifically to TAR RNA, to the average person, the selectivity of the Tat peptide is surprising because at a physiological pH (7.4), all the residues of the peptide would be positively charged. Hence, one would expect that such a positively-charged peptide would bind indiscriminately to any RNA sequence, as the RNA phosphate backbones are negatively-charged.

To better understand the electrostatic contribution of Tat-TAR binding, a salt-effect experiment was carried out. The concentration of salt, in this case NaCl, was varied and the association constant (K_a) was measured at each concentration, yielding the following plot.

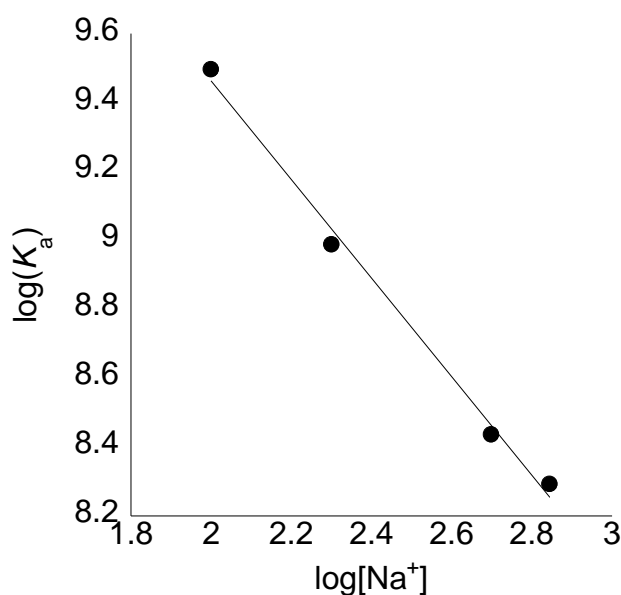


Figure 13. Logarithmic plot of K_a against the concentration of Na cations in the solution. [Probe] = 25 nM, [RNA] = 0 – 125 nM, pH 7.4, 25°C.

The slope of the curve (SK) can be substituted into the following equations to obtain the total observable change in free energy (ΔG_{obs}), the polyelectrolytic contribution (ΔG_{pe}) and the non-polyelectrolytic contribution (ΔG_t). ΔG_{pe} comes from the release of counterions into the solution upon complex formation. ΔG_t includes at least the following terms: $\Delta G_{conf} + \Delta G_{t+r} + \Delta G_{hyd} + \Delta G_{mob}$, where ΔG_{conf} is the free energy contribution from the conformational changes upon complex formation, ΔG_{t+r} is the free energy cost

resulting translational and rotational degrees of freedom upon complex formation, ΔG_{hyd} refers to the free energy relating to the hydrophobic interactions upon complex formation and ΔG_{mol} accounts for the formation of non-covalent molecular interactions (Chaires, 1997).

$$\Delta G_{pe} = (-SK)RT \ln[Na^+]$$

$$\Delta G_{obs} = -RT \ln K_{11} = \Delta G_{pe} + \Delta G_t$$

We compared our obtained free energy values with that of a previously reported wild-type Tat peptide (15-mer) (Suryawanshi, Sabharwal & Maiti, 2010). Although the salt concentrations are not the same, and thus cannot be directly compared, it is worth noting that LUPI binds to TAR RNA with a lower ΔG_{obs} (-12.0 kcal / mol), indicating that LUPI-TAR RNA is a more energetically favourable reaction compared to wild-type Tat peptide-TAR RNA. This is despite the salt concentration of the Tat-TAR reaction being lower than that of LUPI-TAR, as a higher salt concentration interferes with electrostatically-driven binding, and thus we would expect that the ΔG_{pe} of LUPI at [NaCl] = 80 mM would be even lower than what it is, contributing to an even lower ΔG_{obs} .

Nevertheless, for both reactions, most of the change in free energy clearly results from the non-polyelectrolytic contribution, ΔG_t , and that the binding is mostly driven by non-electrostatic interactions between the peptides and the RNA, which explains the specificity between the Tat-TAR interaction.

Table 4. Comparing free energy values between TAR RNA and either LUPI or a wild-type Tat peptide (15-mer).

	[NaCl] / mM	ΔG_{pe} (kcal/mol)	ΔG_t (kcal/mol)	ΔG_{obs} (kcal/mol)
LUPI	110	-1.84	-10.2	-12.0
Tat Peptide (15-mer)	80	-1.92	-7.94	-9.86

9.1.2 Salt Effect Experiments of LUPI with TAR RNA and Bac rRNA

We were also interested in seeing how the free energy values would change for a non-target RNA, and thus we chose to investigate the salt effect on LUPI-Bac rRNA binding and to compare its free energy values with that of LUPI-TAR.

The LUPI-Bac rRNA K_a values are all lower than that of LUPI-TAR, showing that the LUPI-Bac rRNA binding is not as strong. Moreover, the slope of LUPI-Bac rRNA is slightly steeper than that of LUPI-TAR RNA, indicating that LUPI-Bac rRNA is more affected by the varying salt concentration, which implies that the binding is more electrostatically-driven, which is further confirmed by its lower ΔG_{pe} value compared to LUPI-TAR.

This shows that although LUPI can bind to other non-target RNAs, that binding is less strong and more non-specific (more electrostatically-driven).

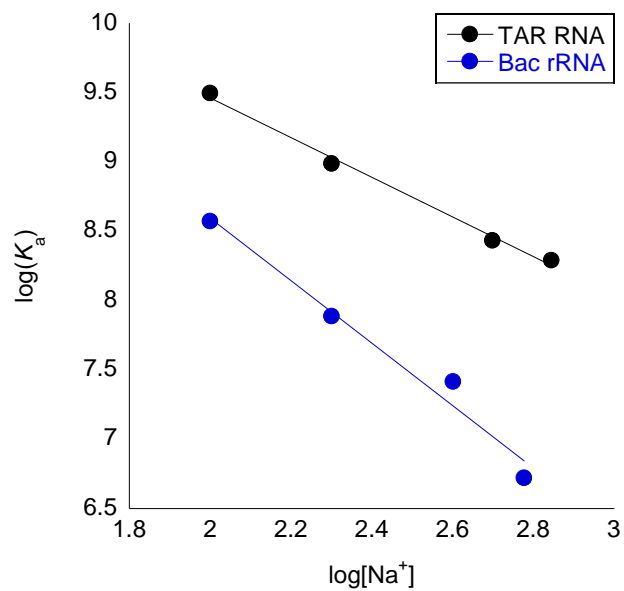


Figure 14. Logarithmic plot of K_a of LUPI-TAR and LUPI-Bac rRNA against the concentration of Na cations in the solution. [Probe] = 25 nM, [RNA] = 0 – 125 nM, pH 7.4, 25°C.

Table 5. Comparing free energy values between LUPI –TAR RNA and LUPI – Bac rRNA.

	[NaCl] / mM	ΔG_{pe} (kcal/mol)	ΔG_t (kcal/mol)	ΔG_{obs} (kcal/mol)
TAR RNA	110	-1.84	-10.2	-12.0
Bac rRNA	110	-2.89	-8.02	-10.9

10. Mock FID Assay

Finally, LUPI was used in a mock FID assay with five test compounds – mitoxantrone, DPQ, neomycin, acridine mutagen ICR 191 and a recombinant Tat protein.

10.1 Introduction to the Test Compounds

The five test compounds chosen had all previously been reported to bind to TAR RNA with varying binding affinities.

Neomycin, an aminoglycoside, has been widely tested as an inhibitor between the Tat-TAR interaction and is the standard mock test compound (Mei et al., 1995; Wang, Huber, Cui, Czarnik & Mei, 1998; Bradrick & Marino, 2004).

Mitoxantrone was first identified through a virtual screen, and then experimentally validated to show that it binds to TAR RNA and can inhibit a fluorescent Tat peptide (Stelzer et al., 2011). Mitoxantrone was not previously known to bind to TAR RNA, and thus these results show how computational screening methods can be useful to finding new drugs.

Acridine Mutagen ICR 191 itself was used as a fluorescent indicator in an FID assay that had TAR RNA as its dsRNA target (Qi, Zhang, He, Huo & Zhang, 2017). As a fluorescent molecule, it would be interesting to see if its fluorescence would hinder the FID assay with LUPI.

DPQ has been shown to inhibit the cellular activity of the influenza A virus promoter region (Lee et al., 2014). However, it was also evaluated for its nonspecific binding to the TAR stem-loop as an example of how small molecule RNA binders often show promiscuous binding (Kelly et al., 2020). It was included as a test compound to see if such a non-selective RNA binder would be able to displace LUPI.

We wanted a test compound that could match that of LUPI's binding affinity ($K_d = 1.0$ nM). This would help us to evaluate the effect of binding affinity on LUPI displacement. However, we could not find a small molecule that had such a low binding affinity and was easily obtainable. Hence, we decided to use the wild-type Tat protein itself as its K_d value was comparable to that of LUPI.

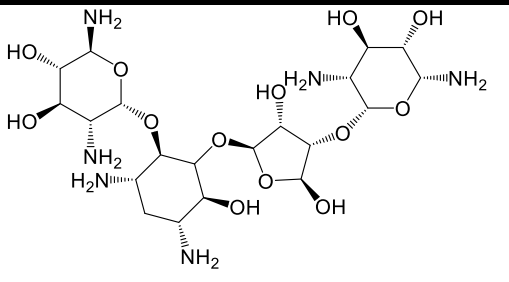
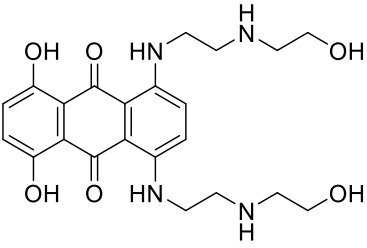
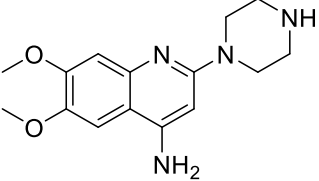
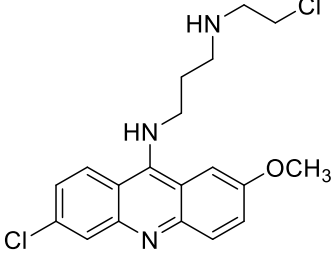
Compound	Chemical Structure	K_d
Neomycin		130 nM
Mitoxantrone		55 nM
DPQ		$K_i = 43 \pm 4 \mu\text{M}^*$
Acridine Mutagen ICR 191		~ 100 nM
Tat Protein (HIV-1 TAT Clade-B Recombinant)	(Amino acid sequence) MEPVDPRLEP WKHPGSQPKT ACTNCYCKKC CFHCQVCFIT KALGISYGRK KRRQRRRPPQ GSQTHQVSLK KQPTSQSRGD PTGPKE	2 – 8 nM

Table 6. A summary of all the test compounds used for the mock FID assay with LUPI. * K_i is the inhibition constant and, similarly to K_d , describes binding affinity.

10.2 Optimising the Conditions for the FID Assay

The three main parameters that had to be optimised were the concentration of LUPI and TAR RNA, the incubation time after the addition of the test compounds, and the number of equivalents of the test compounds.

The aim of the optimised parameters was such that the FID assay would use the least amount of reagents (LUPI, TAR RNA and test compounds) but yet still give good discrimination between hit compounds and non-hit compounds. Minimising the amount of reagent used would allow for the FID assay to be cheaper, and thus more widely usable.

10.2.1 Optimising Incubation Time

The incubation time was the first to be determined, as it would not vary significantly even if the concentrations were changed. To optimise the incubation time, firstly, LUPI and TAR had been incubated together for 30 minutes prior to the addition of the Tat protein. Upon the addition of Tat, the sample were irradiated every five minutes and the fluorescence intensity was recorded. The F.I. was normalised such that at 0 min, the normalised F.I. would be 1. The Tat protein was chosen as the test compound to be added because we assumed that it would result in the most displacement of LUPI, which would facilitate the decision-making for the optimised incubation time. A sample without the Tat protein was similarly measured to serve as a control to account for the natural photodegradation of LUPI.

From Fig. 15, the optimised incubation time was determined to be **10 minutes**.

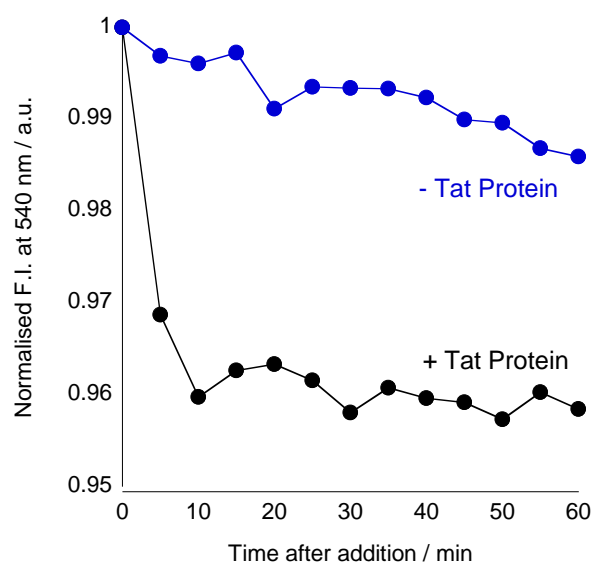


Figure 15. Plot of the normalised F.I. against time to optimise the incubation time for future FID assays. [LUPI] = [TAR RNA] = [Tat Protein] = 25 nM, PBS Buffer, 25°C.

10.2.2 Optimising the Number of Equivalents for the Test Compounds

Many preliminary experiments were done with 5 equivalents of test compounds, as that was the condition used in the FID assay with SPOC (data not shown). However, we quickly realised that we needed to use higher concentrations of test compounds to see a concentration-dependent response. To that end, concentration dependence curves of all the test compounds were measured and plotted in the figure below.

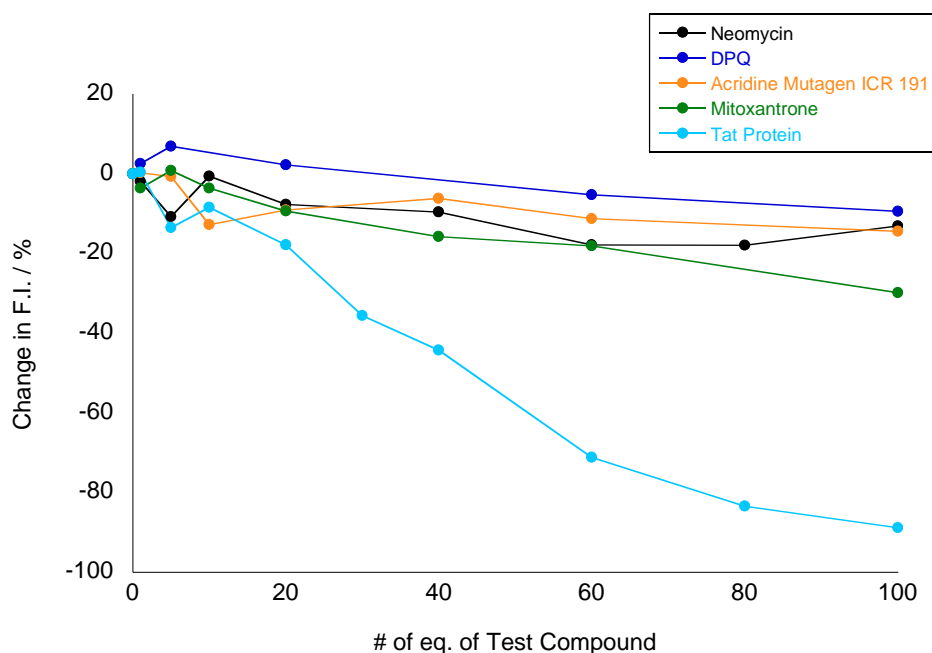


Figure 16. Change in F.I. against the number of equivalents of the test compounds. For the experiment with Tat protein, [LUPI] = [TAR RNA] = 2 nM, [Tat Protein] = 0 – 200 nM. For the rest of the test compounds, [LUPI] = [TAR RNA] = 10 nM, [Test Compound] = 0 – 1000 nM. PBS Buffer, 25°C.

As the objective of the optimisation was to maximise the degree of displacement, especially regarding the Tat protein as the strongest binder amongst all the test compounds, while minimising the amount of test compound used, it was decided that **60 equivalents** of test compound was the optimised amount. This was because 60 eq. resulted in Tat protein causing significantly more displacement compared to the other test compounds, showing a clear discrimination between its binding affinity ($K_d = 2\text{--}8$ nM) and the others, which have a K_d value at least one order higher.

10.2.3 Optimising the Concentrations of LUPI and TAR RNA

The optimised concentration of [LUPI] = [TAR RNA] = 2 nM was determined after considering a few factors.

Firstly, from a technical standpoint, even though the fluorescence spectrophotometer could detect LUPI at 1 nM (Em band = 5 nm, Ex band = 5 nm, Sensitivity = High), the S/N ratio was not favourable (Fig. S9). To counter that, using a sensitivity of “medium” resulted in much better S/N ratios. The lowest concentration that would allow for a sensitivity of “medium” to be used was 2 nM.

Moreover, using lower concentrations for LUPI and TAR RNA would affect the concentration of test compounds used, and as previously mentioned, the lower the amount of test compounds used, the cheaper the overall FID assay would be. Additionally, for fluorescent test compounds, such as Acridine Mutagen ICR 191, a lower concentration would result in less autofluorescence and less interference, allowing for those test compounds to be part of the assay even though they are fluorescent. At 2 nM, the amount of Acridine Mutagen ICR 191 would be 120 nM (60 eq.), which results in its fluorescent response being ~10% of LUPI’s response, which we judged to be low enough to not affect the FID assay (Fig. S10).

Finally, we did an FID assay using FtatRhd at 2 nM and realised that at 2 nM, FtatRhd had a negligible fluorescence response. This clearly showed how using LUPI’s incredible light-up response is advantageous in allowing it to be detected even at such low concentrations, something that previous indicators cannot do.

One drawback of using such a low concentration of LUPI is that the photostability of the fluorescent indicator suffers a little. This is easily circumvented by using fresh LUPI working solutions for each FID assay.

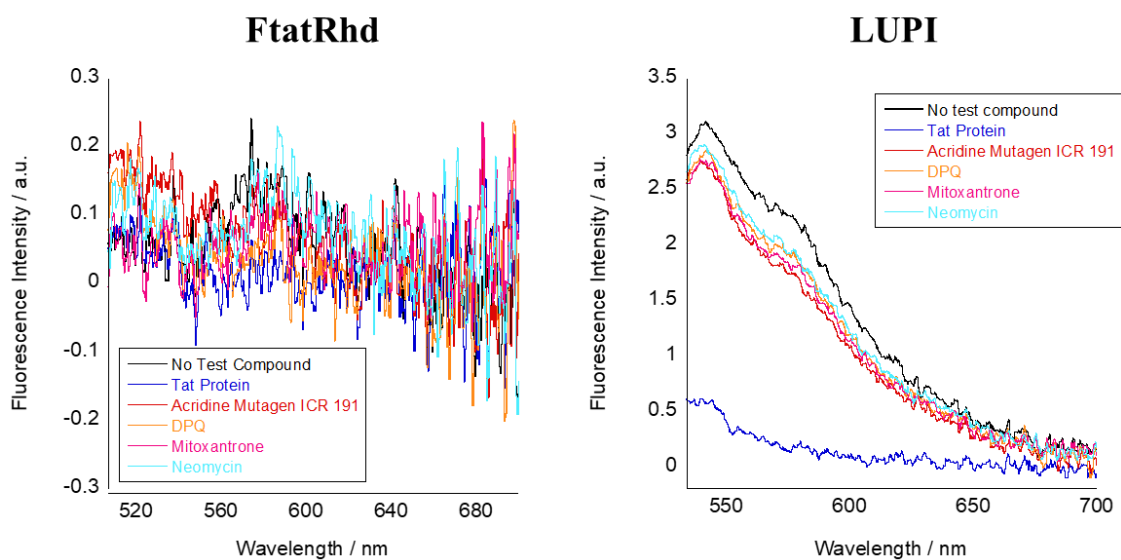


Figure 17. Comparing FID assays done with FtatRhd and LUPI. [Indicator] = [TAR RNA] = 2 nM, [Test Compound] = 120 nM, Excitation wavelength = 497 nm (FtatRhd) or 524 nm (LUPI). Em band = 5 nm, Ex band = 5 nm, Sensitivity = Medium, PBS Buffer, 25°C.

10.3 Optimised Mock FID Assay

With the optimised conditions, we conducted the mock FID assay and triplicated the experiment to get the following figure.

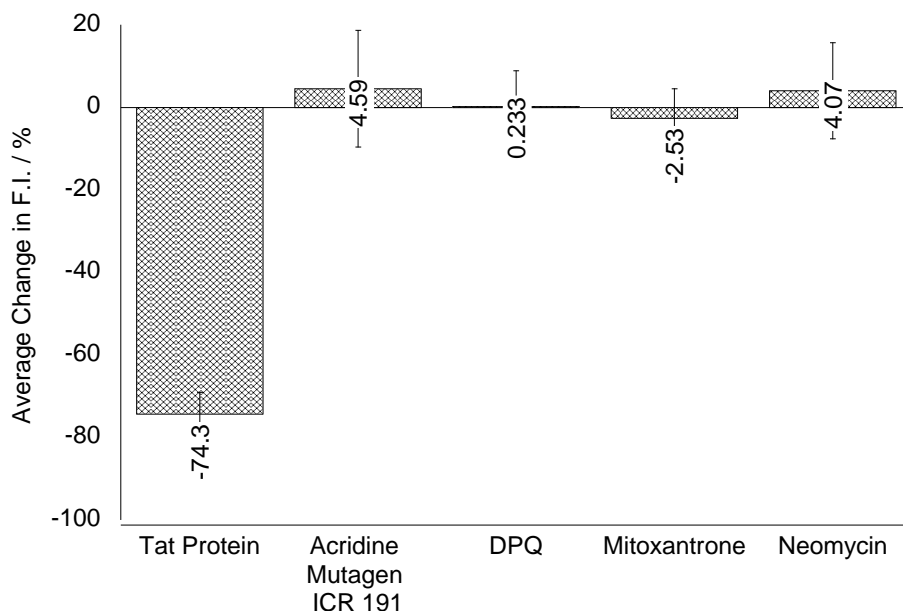


Figure 18. Mock FID assay of LUPI and TAR RNA against five test compounds. [LUPI] = [TAR RNA] = 2 nM, [Test Compound] = 120 nM, Incubation time = 10 min, PBS Buffer, 25°C. Error bars were plotted from the results of three separate experiments.

As seen in Fig.18, under the optimised conditions, the Tat protein resulted in a ~75% displacement while the rest of the test compounds caused negligible displacement. These results were surprising because compounds like neomycin that have been previously used to displace other Tat peptides could not displace LUPI even at 60 eq. This sets LUPI apart as an indicator that can only be displaced by super-strong binders ($K_d < 10$ nM).

This surprising stringency of LUPI as an indicator has its advantages and disadvantages. Some might worry that LUPI is binding too strongly to TAR RNA, and thus might produce false negatives (test compounds that would work well as drugs but could not displace LUPI). Conversely, I would like to propose that this stringency is beneficial. As of 2022, even after more than 30 years of effort from the scientific community, there is still no cure for HIV (Thomas, Ruggiero, Paxton & Pollakis, 2020). Along with new chemical libraries of small molecules that can bind to RNA, perhaps what is also needed is a much stricter FID assay where only small molecules that bind super strongly to TAR RNA can be sieved out and validated. This stringent test would streamline the downstream processes and to facilitate a more efficient screening process.

However, if LUPI's extremely strong binding affinity made it unfavourable for use, it can easily be modified to have weaker binding. For example, LUPI could be shortened from a 9-mer to an 8- or 7-mer, although that would no doubt have repercussions on its selectivity as well. As shown by our experiments with the conventional probe with TO coupled to the terminal of the probe, tuning the binding affinity of LUPI is also possible via changing the position of TO in the probe.

What is clear from this mock FID assay is that LUPI can indeed function as an FID indicator, and especially noteworthy is its ability to sieve out super-strong binders.

11. Conclusion

We have successfully applied the light-up peptide indicator (LUPI) concept to the Tat – TAR model of the HIV-1 virus by showing how thiazole orange (TO) can act as a surrogate amino acid. By substituting a glutamine residue with TO, LUPI was able to exhibit a significant light-up response upon binding to TAR RNA, showing how TO was able to effectively intercalate into the dsRNA. Furthermore, LUPI has an incredibly strong binding affinity with TAR RNA ($K_d = 1.0 \pm 0.6$ nM), comparable to the wild-type Tat protein itself. LUPI maintains its specific binding to TAR RNA compared to two other non-target RNAs and is able to act as an FID indicator under conditions that the previous indicator, FtatRhd, failed to work at.

Moreover, this research shows the importance of internally-incorporating the fluorophore to increase the probe's fluorescence response and binding affinity. Experiments conducted with the terminally-linked conventional probe showed a significant decrease in fluorescence intensity, binding affinity and selectivity compared to LUPI.

Varying the linker length also affects the intercalation of TO into the dsRNA, as our experiments with the Lys-linker probe have shown. Even at the same position, a longer linker might make TO too flexible to effectively intercalate, leading to it being unable to discriminate between Bac rRNA and Influenza vRNA, something that the rest of the probes could do.

LUPI's incredibly low LOD with TAR RNA (18 pM) also shows its potential for other applications. We tried using LUPI to image the nucleolus in MCF7 cells (both live and fixed cells) (Fig. S12), however LUPI was unable to clearly discriminate between the nucleolus and the surrounding nucleic acids. We hope to try other applications in the future.

Matsumoto and his team developed the FtatRhd probe 22 years ago. We hope that in the future, LUPI would be able to widely used as its successor and that through its use, new drugs for the HIV-1 virus will be discovered.

12. Experimental Procedures

Most of the experimental procedures are the same as that of Chapter 2. Only procedures that were conducted differently / chemicals that were specific to this chapter will be listed here.

12.1 General Information

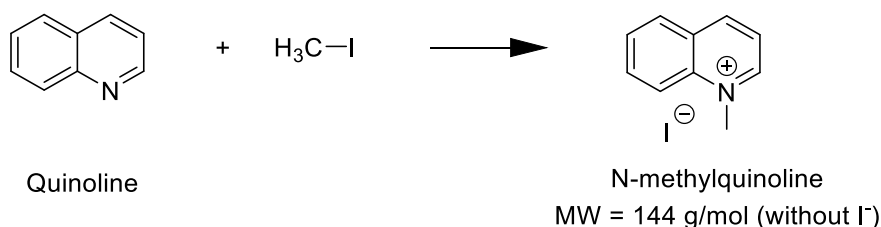
Fmoc amino acids were purchased from

TAR RNA was custom synthesized and HPLC purified (99%) by GeneDesign, Inc. (Osaka, Japan). The concentration of the RNAs was determined from its UV absorbance value at 260 nm and calculated using the sum of the average molar extinction coefficients of its individual nucleobases at 25°C.

FtatRhd was custom-made by HiPep Laboratories (Kyoto, Japan). FtatRhd concentration was measured using UV-Vis spectrometry in MeOH and calculated using $\epsilon_{\text{TAMRA}} = 95,000 \text{ M}^{-1}\text{cm}^{-1}$ (Voss, Fischer, Jung, Wiesmüller & Brock, 2006).

HIV-1 TAT Clade-B Recombinant protein was purchased from ProSpec (Rehovot, Israel). Its concentration was determined using the manufacturer-provided mass of the protein and its theoretical molecular weight (9.79 kDa).

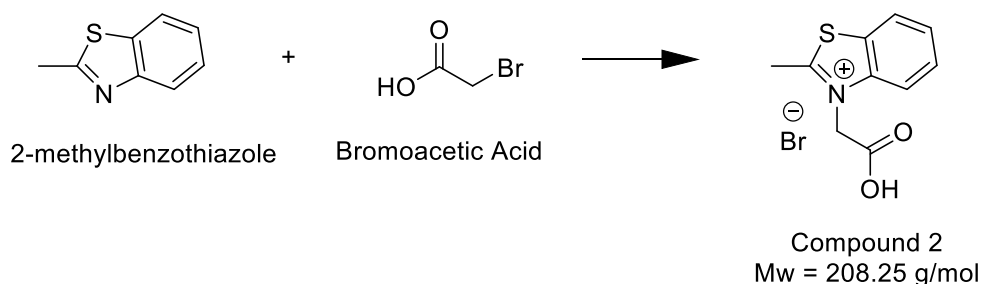
12.2 TO_z-C₁-COOH Synthesis



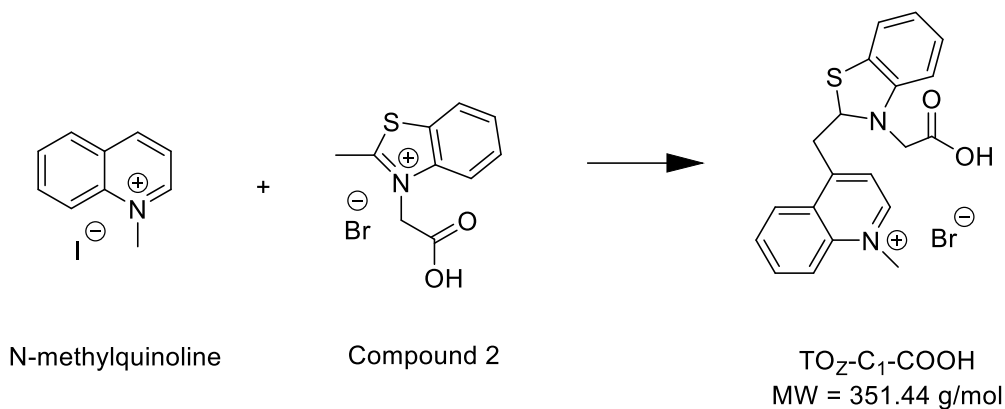
TO_z-C₁-COOH (1): Quinoline (2.01 g) and iodomethane (7.06 g) was added to a 100 mL round bottom flask. After the addition of dry 1,4-dioxane (34 mL), the mixture was refluxed under an N₂ environment for 1h. After cooling to room temperature, the yellow solid (N-methylquinoline) was filtered out and characterised by ESI-MS and ¹H NMR.

ESI-MS: Peak at $m/z = 144$, [M]⁺ without I⁻

¹H NMR (500 MHz, MeOD): δ (ppm) = 9.3 (d, 1H, $J = 6 \text{ Hz}$), 9.14 (d, 1H, $J = 8 \text{ Hz}$), 8.44 (d, 1H, $J = 9 \text{ Hz}$), 8.36 (d, 1H, $J = 9 \text{ Hz}$), 8.23 (dt, 1H, $J = 7.5 \text{ Hz}$), 7.98 – 8.02 (m, 2H), 4.63 (s, 3H).



TO_z-C₁-COOH (2): Bromoacetic acid (7.20 g) and 2-methylbenzothiazole (7.00 g) were added to a dry 100 mL round bottom flask and refluxed for 1.5 h under N₂ gas. The resultant solid was dried in a desiccator before ~60 mL of MeOH was added to it. Diethyl ether (40 mL) was added dropwise to the mixture and a very pale pink solid was filtered out, which was Compound 2, which was confirmed with ESI-MS. (Peak at m/z = 208, [M]⁺ without Br⁻). Compound 2 was used without further purification.



TO_z-C₁-COOH (3): To a dry round bottom flask was added 1.88 g (7.0 mmol) of N-methylquinolinium iodide and compound 2 (2.38 g, 1.2 eq.) before dry DCM (30 mL) was added, The resultant yellow-orange mixture was stirred for 5 min before turning red upon the addition of 2.10 g of Et₃N. This mixture was stirred at r.t. for 24 h in an inert environment to give a deep red residue, which was dissolved in EtOH (40 mL) and recrystallised with diethyl ether (12.5 mL). The precipitate was collected using vacuum filtration and confirmed with ¹H NMR.

¹H NMR data (500 MHz, DMSO-D₆): δ (ppm) = 8.59 (dd, 2H), 8.01 – 8.03 (m, 3H), 7.30 – 7.69 (m, 5H), 6.84 (s, 1H), 5.20 (s, 2H), 4.16 (s, 3H).

12.3 Quantifying LUPI (TO_Z)

The probe carrying TO_Z was quantified using the molar extinction coefficient of TO_Z-C₁-COOH, which was obtained through a calibration curve (Fig. S5). A stock solution of TO_Z-C₁-COOH was made and dissolved into a series of concentrations before the absorbance of each sample was measured.

From the slope of the curve, $\epsilon_{\text{TO}_Z\text{-C}_1\text{-COOH}} = 59,166 \text{ M}^{-1} \text{ cm}^{-1}$

12.3 Absolute Fluorescence Quantum Yield Measurement

The Hamamatsu Absolute PH Quantum Yield Spectrometer C11347 Quantaurus – QY machine was used for the absolute fluorescence quantum yield measurements.

The excitation wavelength was set to 510 nm, which is the absorbance peak of LUPI.

[LUPI] = 300 nM, [TAR RNA] = 2250 nM (7.5 eq.) in order to ensure that all of LUPI is hybridised with TAR RNA, to maximise the fluorescence intensity.

12.4 FID Assay protocol

The FID assays were carried out using the following protocol.

The TAR RNA was first re-annealed to ensure that it had the correct stem-loop structure, after which LUPI was added to each sample and allowed to hybridise for at least 30 minutes. After 30 minutes, the test compounds were added and allowed to incubate for 10 minutes before LUPI's fluorescent response was measured immediately after.

12.5 Fluorescence Imaging in Live Cells

MCF7 cells were grown in RPMI 1640 media supplemented with 10% fetal bovine serum and 2% penicillin/streptomycin at 37°C in a 5% CO₂ S3 incubator. For the fluorescence imaging experiments, the cells were seeded in an 8-chambered glass plate (Iwaki, Tokyo, Japan) at a density of about 5.0×10³ cells/well and maintained for 24 h. The cells were incubated in media containing appropriate concentration of probe for 1 h at 37°C in a 5% CO₂ incubator. After washing with HBSS buffer twice, the cells were imaged in a HBSS buffer using a Deltavision Elite microscopy system (GE Healthcare Japan, Tokyo, Japan).

The following filter set was used: FITC filter set (Ex 475/28; Em 545/48) for LUPI. The obtained images were processed with the softWoRx software.

12.6 Fluorescence imaging in fixed cells

MCF-7 cells were grown in RPMI 1640 media supplemented with 10% fetal bovine serum and 2% penicillin/streptomycin at 37°C in a 5% CO₂ incubator. For the fluorescence imaging experiments, MCF-7 cells were seeded in a 8-chambered glass plate (Iwaki, Tokyo, Japan) at a density of about 5.0×10³ cells/well and maintained for 24 h. The cells were fixed in pre-chilled methanol at -20°C for 1 min. The cell membrane was permeabilized with 1% Triton X-100 in PBS for 2 min at room temperature. After rinsing with PBS twice, The cells were incubated in PBS containing appropriate concentration of probes for 1 h at 37°C in a 5% CO₂ incubator. After washing with PBS buffer twice, the cells were imaged in a PBS using a Deltavision Elite microscopy system (Cytiva, Tokyo, Japan). The following filter set was used: FITC filter set (Ex 475/28; Em 545/48) for LUPI. The obtained images were processed with the softWoRx software.

13. Supporting Information

HPLC Chromatograms and MALDI-TOF Mass spectra of synthesised probes

LUPI (TO_Q)

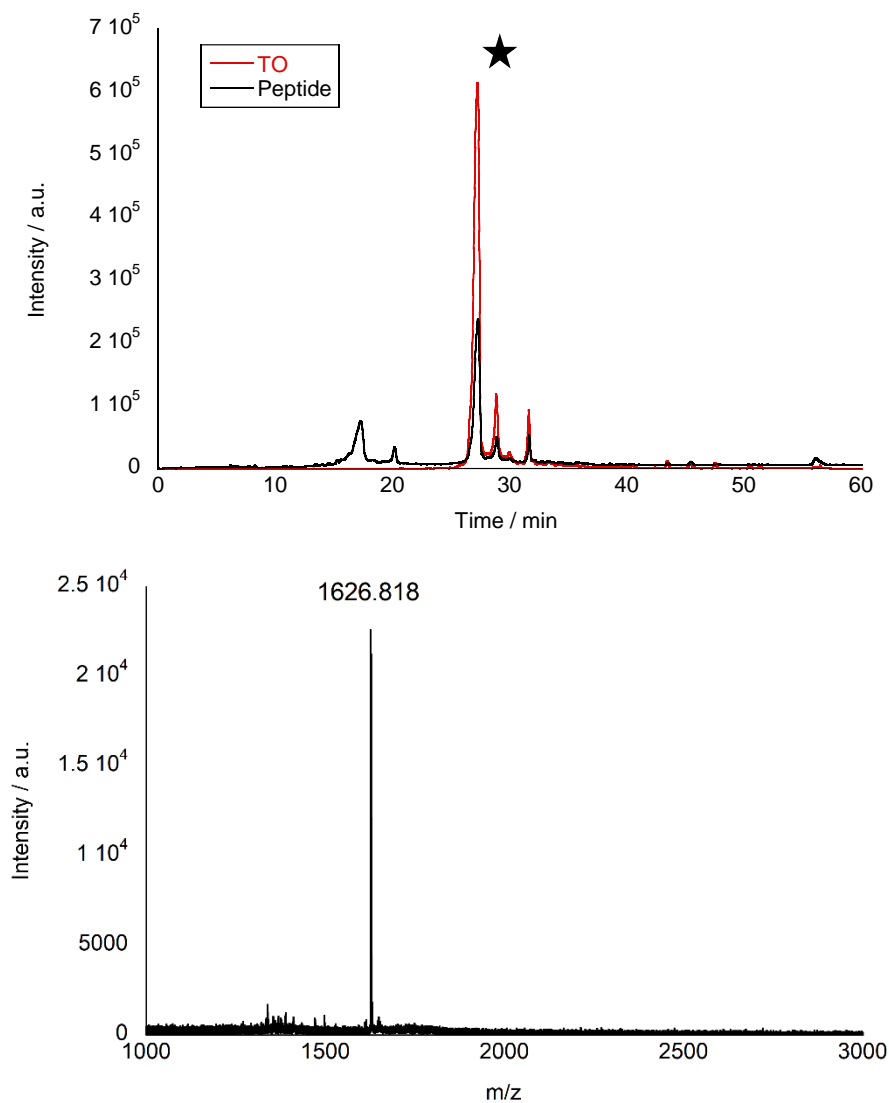


Figure S1. (Top) HPLC chromatogram of LUPI (TO_Q). The peak corresponding to LUPI (TO_Q) is marked with a star. **(Bottom)** MALDI-TOF spectrum of LUPI (TO_Q) after HPLC purification. Matrix: CHCA.

LUPI (TO_Z)

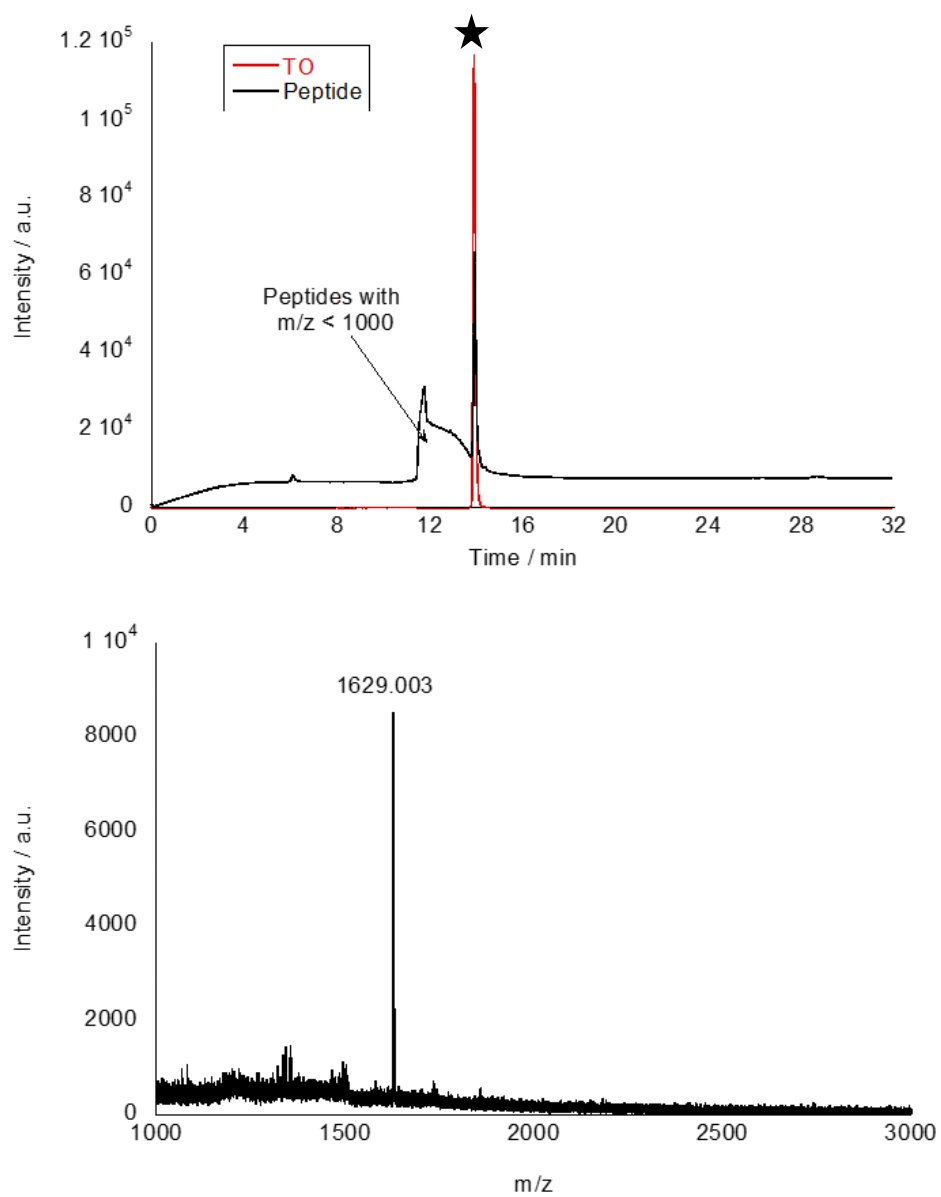


Figure S2. (Top) HPLC chromatogram of LUPI (TO_Z). The peak corresponding to LUPI (TO_Z) is marked with a star. **(Bottom)** MALDI-TOF spectrum of LUPI (TO_Z) after HPLC purification. Matrix: CHCA.

Conventional Probe

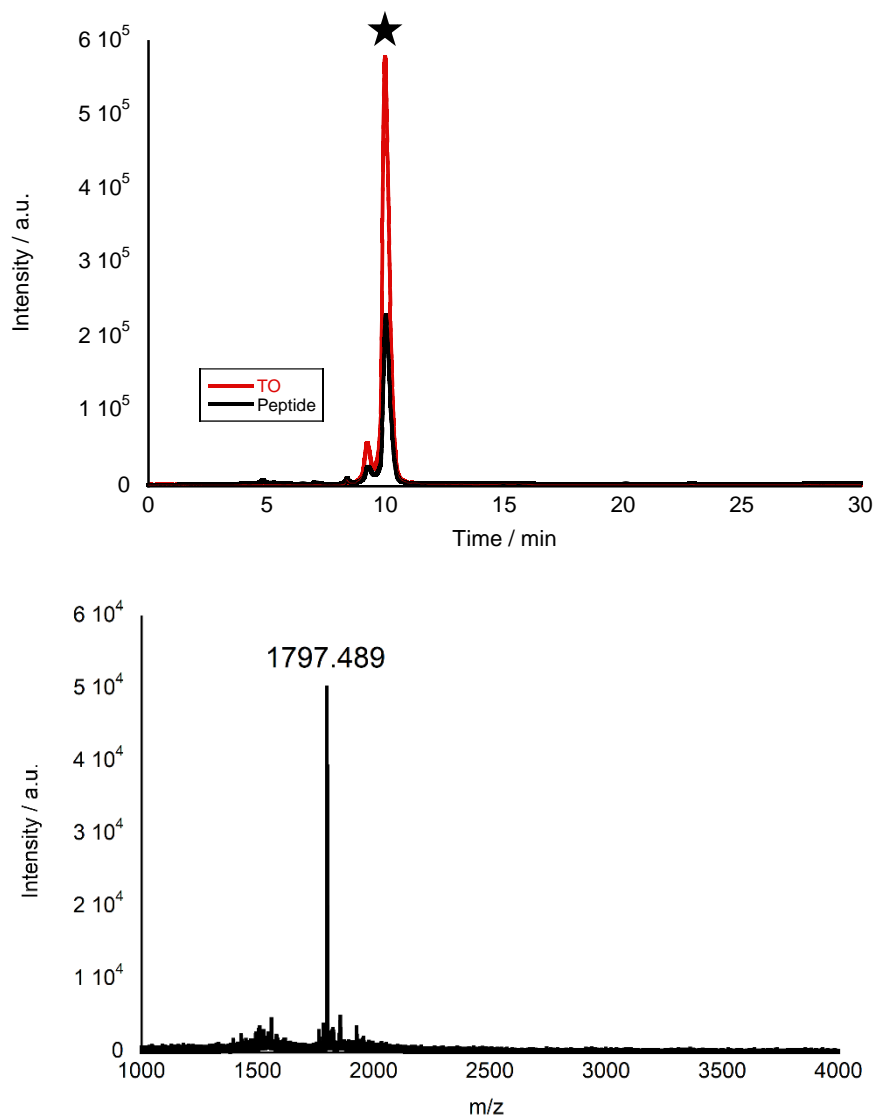


Figure S3. (Top) HPLC chromatogram of the conventional probe. The peak corresponding to the conventional probe is marked with a star. (Bottom) MALDI-TOF spectrum of the conventional probe after HPLC purification. Matrix: CHCA.

Lys-Linker Probe

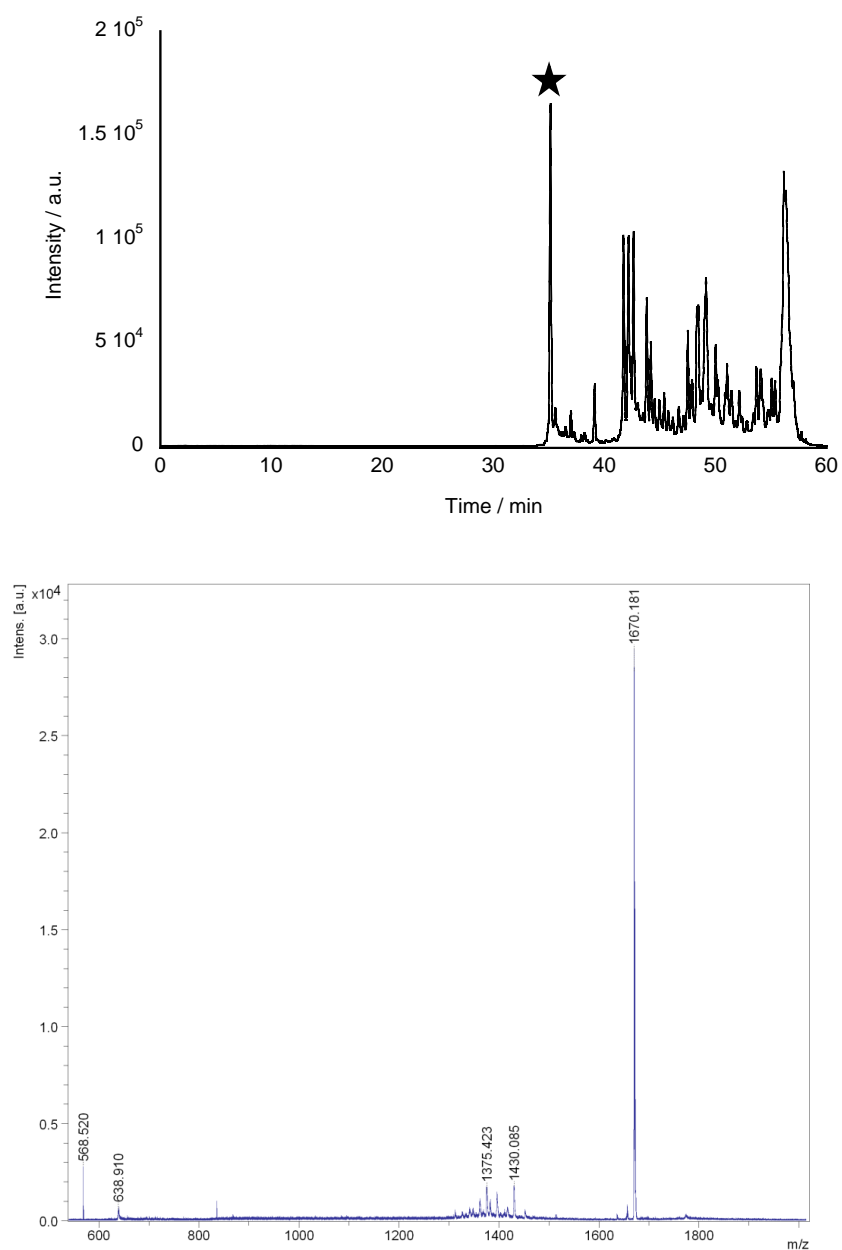


Figure S4. (Top) HPLC chromatogram of the Lys-linker probe. The peak corresponding to the Lys-linker probe is marked with a star. (Bottom) MALDI-TOF spectrum of the Lys-linker probe after HPLC purification. Matrix: CHCA.

Probe	Predicted m/z	Actual m/z
LUPI (TO _Q)	1626.96	1626.818
LUPI (TO _Z)	1626.96	1629.003
Conventional Probe	1797.07	1797.489
Lys-linker Probe	1669.01	1670.181

Table S1. Table of predicted and actual m/z values of the synthesised probes.

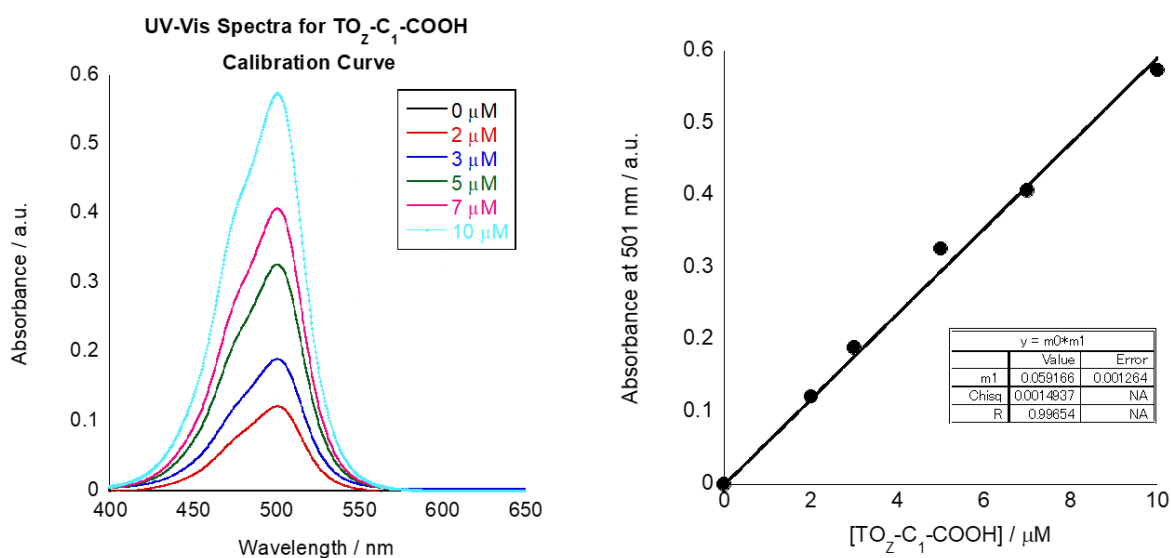


Figure S5. UV-Vis spectra and calibration curve of TO_Z-C₁-COOH. From the slope of the calibration curve, $\epsilon_{\text{TO}_Z\text{-C}_1\text{-COOH}}$ was calculated to be $59,166 \text{ M}^{-1} \text{ cm}^{-1}$.

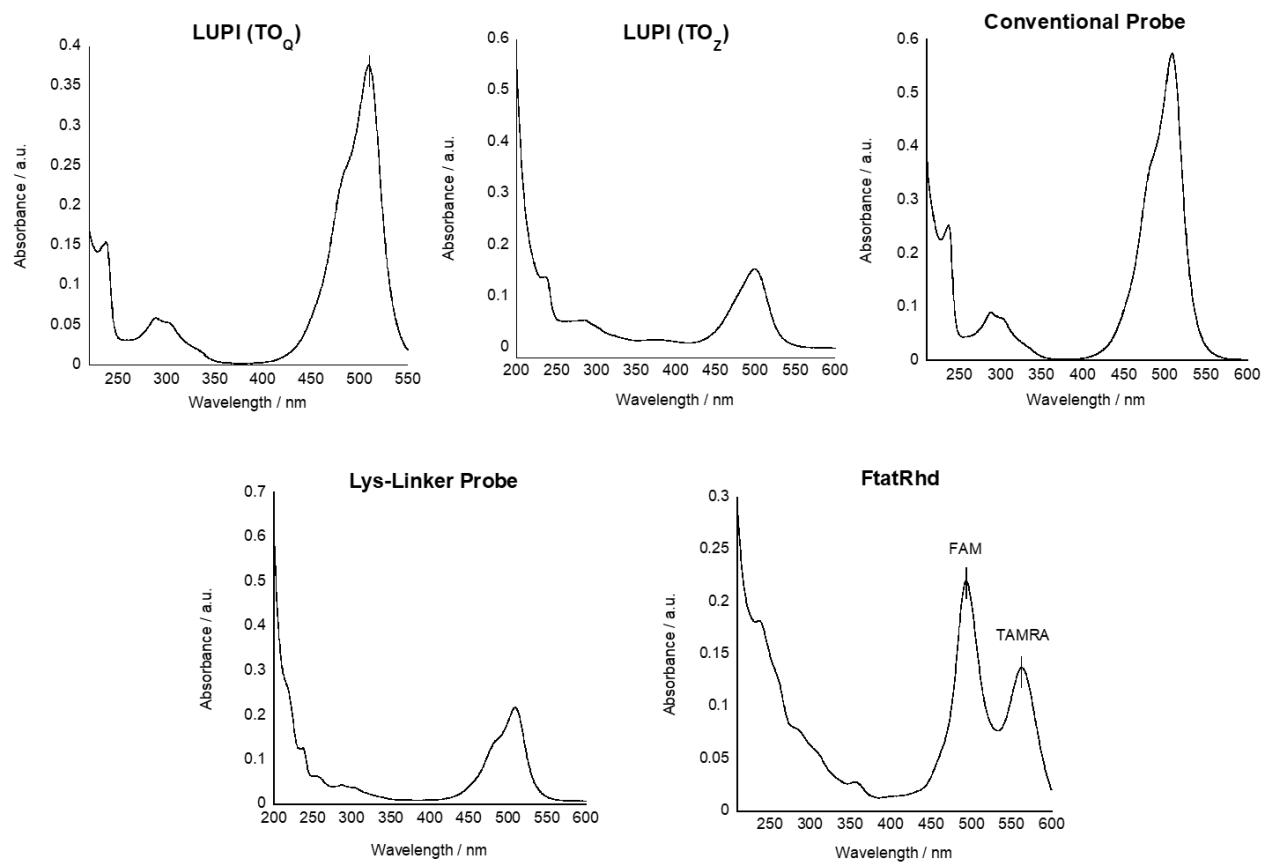


Figure S6. UV-Vis absorbance spectra of all five probes in MilliQ.

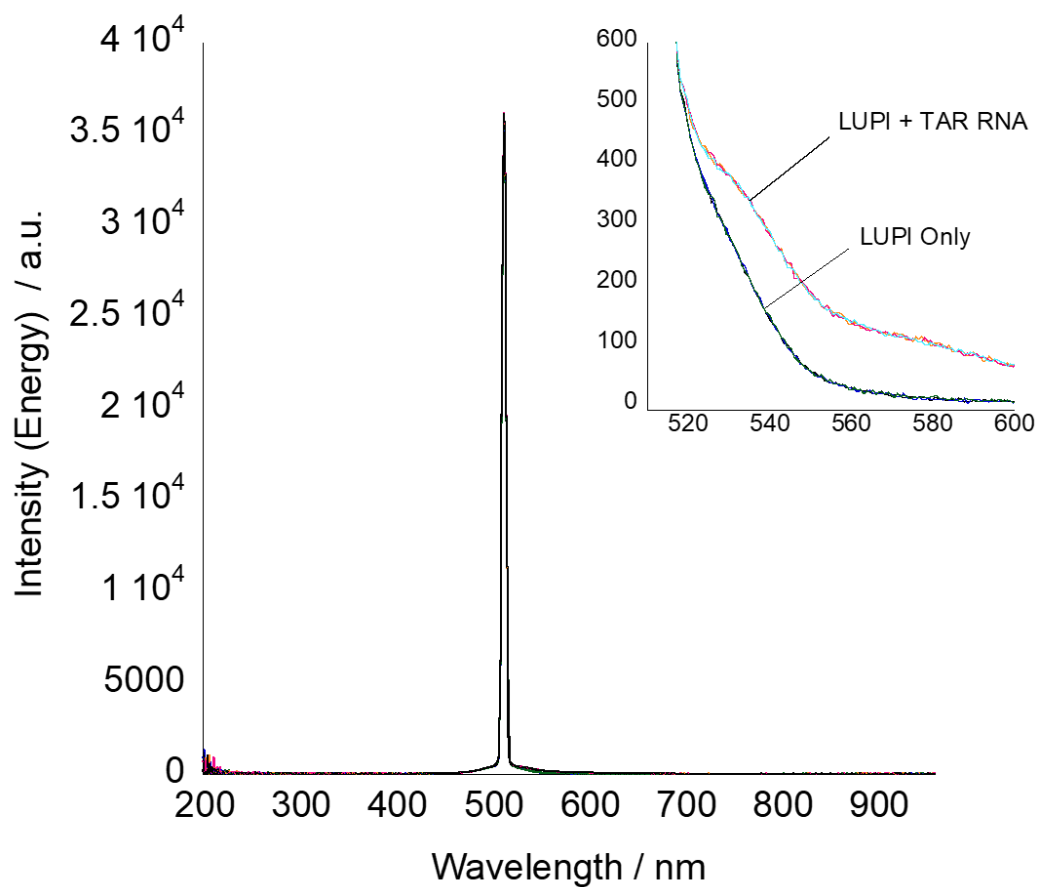


Figure S7. Fluorescence spectrum of LUPI and TAR RNA to measure the absolute fluorescence quantum yield of LUPI with and without TAR RNA. [LUPI] = 300 nM, [TAR RNA] = 2250 nM, Excitation wavelength = 510 nm, PBS Buffer. Each sample was measured three times and the average value was used as its absolute fluorescence quantum yield.

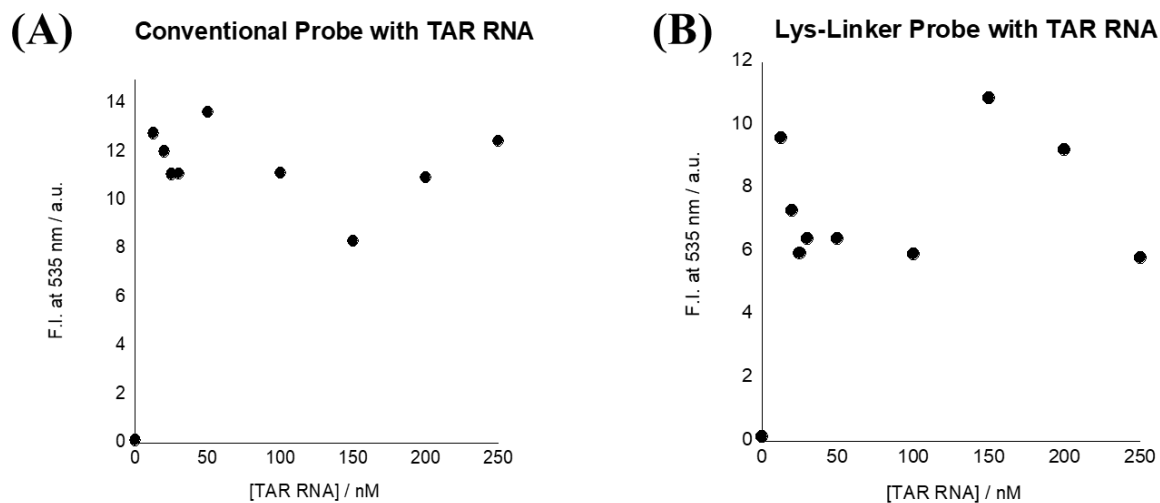


Figure S8. Concentration dependence curves for **(A)** conventional probe with TAR RNA and **(B)** Lys-linker probe with TAR RNA. [Probe] = 25 nM, [TAR RNA] = 0 - 250 nM, Excitation wavelength = 521 nm, PBS Buffer, 25°C. We were unable to fit a binding isotherm to these plots, indicating that the K_d values of these two probes were much higher than LUPI's, and thus the conditions chosen to measure LUPI's K_d value were not appropriate for them.

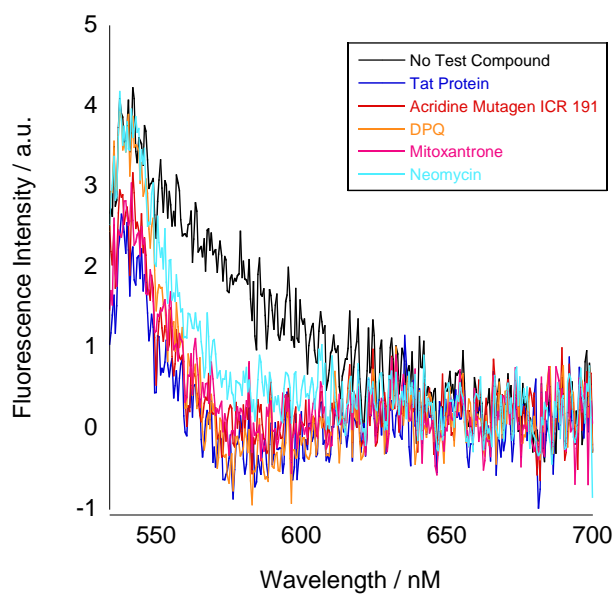


Figure S9. FID assay done when [LUPI] = [TAR RNA] = 1 nM. The S/N ratio was unfavourable, even though displacement could be seen. [Test Compound] = 10 nM, Em band = 5 nm, Ex band = 5 nm, Sensitivity = High, PBS Buffer, 25°C.

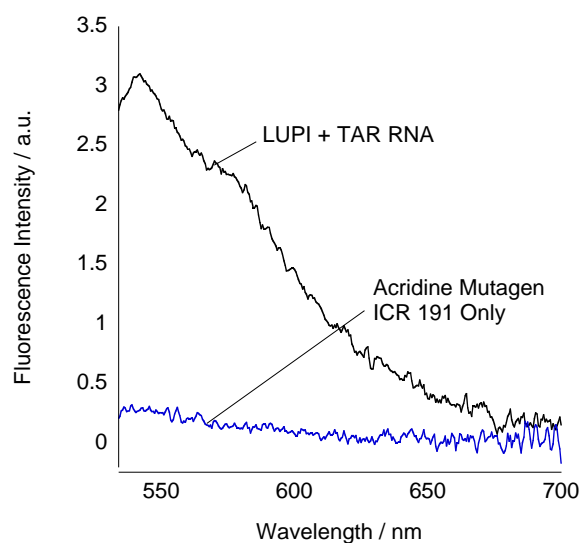


Figure S10. Fluorescence spectrum comparing Acridine Mutagen ICR 191's autofluorescence with LUPI's light-up response. [LUPI] = [TAR RNA] = 2 nM, [Acridine Mutagen ICR 191] = 120 nM. Em band = 5 nm, Ex band = 5 nm, Sensitivity = Medium, Excitation wavelength = 524 nm. PBS Buffer, 25°C.

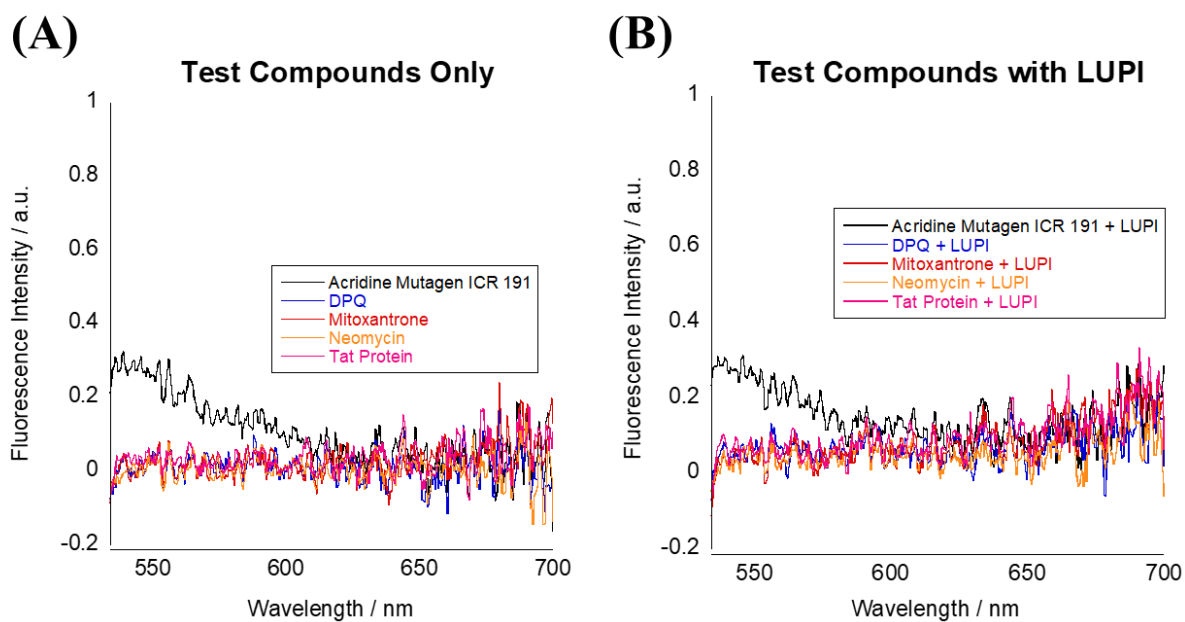
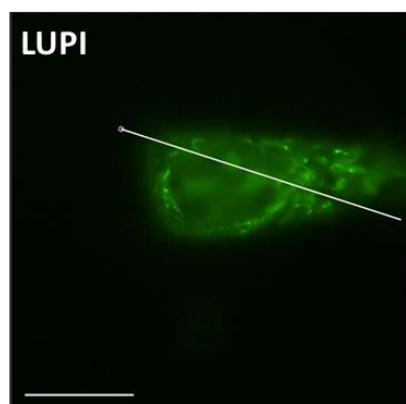


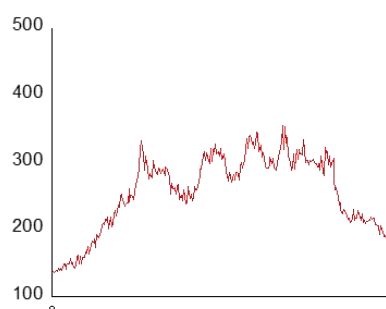
Figure S11. Fluorescence spectrum measuring the autofluorescence of (A) all the test compounds alone or (B) with LUPI. [LUPI] = 2 nM, [Test Compound] = 120 nM. Em band = 5 nm, Ex band = 5 nm, Sensitivity = Medium, Excitation wavelength = 524 nm. PBS Buffer, 25°C.

Only Acridine Mutagen ICR 191 emits fluorescence under these conditions. All other test compounds are non-fluorescent.

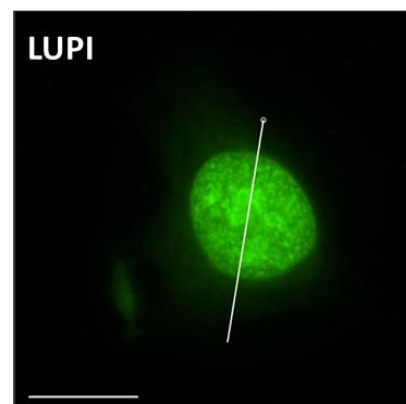
(A) Live cell imaging



[Probe] = 1 μ M, Incubation = 1 h
%T 50, Ex 0.1
Scale bar indicates 15 μ m



(B) Fixed cell imaging



[Probe] = 1 μ M, Incubation = 1 h
%T 5, Ex 0.1
Scale bar indicates 15 μ m

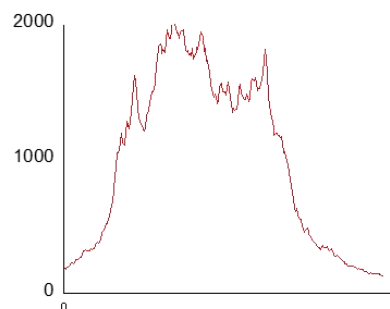


Figure S12. LUPI was used as an imaging probe for the nucleolus in MCF7 (A) live cells and (B) fixed cells. In the live cell image, the nucleolus can vaguely be made out, however its fluorescence intensity was low. In the case of the fixed cells, the fluorescence intensity is much higher, but LUPI could not discriminate between the nucleolus and the surround nucleic acids. Overall, LUPI is unfortunately not suitable as a probe for the nucleolus.

14. References

- Aboul-ela, F., Karn, J., & Varani, G. (1995). The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *Journal Of Molecular Biology*, 253(2), 313-332.
- Aboul-ela, F. (1996). Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Research*, 24(20), 3974-3981.
- Anand, K., Schulte, A., Vogel-Bachmayr, K., Scheffzek, K., & Geyer, M. (2008). Structural insights into the Cyclin T1–Tat–TAR RNA transcription activation complex from EIAV. *Nature Structural & Molecular Biology*, 15(12), 1287-1292.
- Barré-Sinoussi, F., Chermann, J., Rey, F., Nugeyre, M., Chamaret, S., & Gruest, J. et al. (1983). Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science*, 220(4599), 868-871.
- Bradrick, T., & Marino, J. (2004). Ligand-induced changes in 2-aminopurine fluorescence as a probe for small molecule binding to HIV-1 TAR RNA. *RNA*, 10(9), 1459-1468.
- Brannon, J., & Magde, D. (1978). Absolute quantum yield determination by thermal blooming. Fluorescein. *The Journal Of Physical Chemistry*, 82(6), 705-709.
- Brodsky, A., & Williamson, J. (1997). Solution structure of the HIV-2 TAR-argininamide complex. *Journal Of Molecular Biology*, 267(3), 624-639.
- Calnan, B., Tidor, B., Biancalana, S., Hudson, D., & Frankel, A. (1991). Arginine-Mediated RNA Recognition: the Arginine Fork. *Science*, 252(5009), 1167-1171.
- Carpino, L., Shroff, H., Triolo, S., Mansour, E., Wenschuh, H., & Albericio, F. (1993). The 2,2,4,6,7-pentamethylidihydrobenzofuran-5-sulfonyl group (Pbf) as arginine side chain protectant. *Tetrahedron Letters*, 34(49), 7829-7832.
- Carreon, J., Mahon, K., & Kelley, S. (2004). Thiazole Orange–Peptide Conjugates: Sensitivity of DNA Binding to Chemical Structure. *Organic Letters*, 6(4), 517-519.
- Chaires, J. (1997). Energetics of drug–DNA interactions. *Biopolymers*, 44(3), 201-215.
- Chaloin, O., Peter, J., Briand, J., Masquida, B., Desgranges, C., Muller, S., & Hoebeke, J. (2005). The N-terminus of HIV-1 Tat protein is essential for Tat-TAR RNA interaction. *CMLS Cellular And Molecular Life Sciences*, 62(3), 355-361.
- D'Agostino, V., Sighel, D., Zucal, C., Bonomo, I., Micaelli, M., & Lolli, G. et al. (2019). Screening Approaches for Targeting Ribonucleoprotein Complexes: A New Dimension for Drug Discovery. *SLAS Discovery*, 24(3), 314-331.

- Debaisieux, S., Rayne, F., Yezid, H., & Beaumelle, B. (2011). The Ins and Outs of HIV-1 Tat. *Traffic*, 13(3), 355-363.
- Delling, U., Roy, S., Sumner-Smith, M., Barnett, R., Reid, L., Rosen, C., & Sonenberg, N. (1991). The number of positively charged amino acids in the basic domain of Tat is critical for trans-activation and complex formation with TAR RNA. *Proceedings Of the National Academy Of Sciences*, 88(14), 6234-6238.
- Dingwall, C., Ernberg, I., Gait, M., Green, S., Heaphy, S., & Karn, J. et al. (1990). HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. *The EMBO Journal*, 9(12), 4145-4153.
- Dolicka, D., Sobolewski, C., Correia de Sousa, M., Gjorgjieva, M., & Foti, M. (2020). mRNA Post-Transcriptional Regulation by AU-Rich Element-Binding Proteins in Liver Inflammation and Cancer. *International Journal Of Molecular Sciences*, 21(18), 6648.
- Edwards, T., Robinson, B., & Sigurdsson, S. (2005). Identification of Amino Acids that Promote Specific and Rigid TAR RNA-Tat Protein Complex Formation. *Chemistry & Biology*, 12(3), 329-337.
- Galarneau, A., & Richard, S. (2005). Target RNA motif and target mRNAs of the Quaking STAR protein. *Nature Structural & Molecular Biology*, 12(8), 691-698.
- Gerstberger, S., Hafner, M., & Tuschl, T. (2014). A census of human RNA-binding proteins. *Nature Reviews Genetics*, 15(12), 829-845.
- Greenbaum, N. (1996). How Tat targets TAR: structure of the BIV peptide-RNA complex. *Structure*, 4(1), 5-9.
- Guo, Y., Wang, W., Sun, Y., Ma, C., Wang, X., & Wang, X. et al. (2016). Crystal Structure of the Core Region of Hantavirus Nucleocapsid Protein Reveals the Mechanism for Ribonucleoprotein Complex Formation. *Journal Of Virology*, 90(2), 1048-1061.
- He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., & Burlingame, A. et al. (2010). HIV-1 Tat and Host AFF4 Recruit Two Transcription Elongation Factors into a Bifunctional Complex for Coordinated Activation of HIV-1 Transcription. *Molecular Cell*, 38(3), 428-438.
- Jeong, H., Choi, S., Kim, H., Park, J., Park, H., & Park, S. et al. (2013). Fluorescent peptide indicator displacement assay for monitoring interactions between RNA and RNA binding proteins. *Molecular Biosystems*, 9(5), 948-951.
- Karn, J. (1999). Tackling tat. *Journal Of Molecular Biology*, 293(2), 235-254.

- Kelly, M., Chu, C., Shi, H., Ganser, L., Bogerd, H., & Huynh, K. et al. (2020). Understanding the characteristics of nonspecific binding of drug-like compounds to canonical stem-loop RNAs and their implications for functional cellular assays. *RNA*, 27(1), 12-26.
- Lee, M., Bottini, A., Kim, M., Bardaro, M., Zhang, Z., & Pellecchia, M. et al. (2014). A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication. *Chem. Commun.*, 50(3), 368-370.
- Leroux, C., Cadoré, J., & Montelaro, R. (2004). Equine Infectious Anemia Virus (EIAV): what has HIV's country cousin got to tell us?. *Veterinary Research*, 35(4), 485-512.
- Matsumoto, C., Hamasaki, K., Mihara, H., & Ueno, A. (2000). A high-throughput screening utilizing intramolecular fluorescence resonance energy transfer for the discovery of the molecules that bind HIV-1 TAR RNA specifically. *Bioorganic & Medicinal Chemistry Letters*, 10(16), 1857-1861.
- Mei, H., Galan, A., Halim, N., Mack, D., Moreland, D., & Sanders, K. et al. (1995). Inhibition of an HIV-1 Tat-derived peptide binding to TAR RNA by aminoglycoside antibiotics. *Bioorganic & Medicinal Chemistry Letters*, 5(22), 2755-2760.
- Metzger, A., Schindler, T., Willbold, D., Kraft, M., Steegborn, C., & Volkmann, A. et al. (1996). Structural rearrangements on HIV-1 Tat (32-72) TAR complex formation. *FEBS Letters*, 384(3), 255-259.
- Nathans, R., Cao, H., Sharova, N., Ali, A., Sharkey, M., & Stranska, R. et al. (2008). Small-molecule inhibition of HIV-1 Vif. *Nature Biotechnology*, 26(10), 1187-1192.
- Patwardhan, N., Cai, Z., Newson, C., & Hargrove, A. (2019). Fluorescent peptide displacement as a general assay for screening small molecule libraries against RNA. *Organic & Biomolecular Chemistry*, 17(7), 1778-1786.
- Peterlin, B., & Price, D. (2006). Controlling the Elongation Phase of Transcription with P-TEFb. *Molecular Cell*, 23(3), 297-305.
- Pugliese, A., Vidotto, V., Beltramo, T., Petrini, S., & Torre, D. (2005). A review of HIV-1 Tat protein biological effects. *Cell Biochemistry And Function*, 23(4), 223-227.
- Puglisi, J., Tan, R., Calnan, B., Frankel, A., & Williamson, J. (1992). Conformation of the TAR RNA-Arginine Complex by NMR Spectroscopy. *Science*, 257(5066), 76-80.
- Qi, L., Zhang, J., He, T., Huo, Y., & Zhang, Z. (2017). Probing interaction of a fluorescent ligand with HIV TAR RNA. *Spectrochimica Acta Part A: Molecular And Biomolecular Spectroscopy*, 173, 93-98.

- Roy, S., Delling, U., Chen, C., Rosen, C., & Sonenberg, N. (1990). A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes & Development*, 4(8), 1365-1373.
- Sato, T., Sato, Y., & Nishizawa, S. (2016). Triplex-Forming Peptide Nucleic Acid Probe Having Thiazole Orange as a Base Surrogate for Fluorescence Sensing of Double-stranded RNA. *Journal Of The American Chemical Society*, 138(30), 9397-9400.
- Sato, T., Sato, Y., & Nishizawa, S. (2017). Optimization of the Alkyl Linker of TO Base Surrogate in Triplex-Forming PNA for Enhanced Binding to Double-Stranded RNA. *Chemistry - A European Journal*, 23(17), 4079-4088.
- Schulze-Gahmen, U., Echeverria, I., Stjepanovic, G., Bai, Y., Lu, H., & Schneidman-Duhovny, D. et al. (2016). Insights into HIV-1 proviral transcription from integrative structure and dynamics of the Tat:AFF4:P-TEFb:TAR complex. *Elife*, 5.
- Shortridge, M., Wille, P., Jones, A., Davidson, A., Bogdanovic, J., & Arts, E. et al. (2018). An ultra-high affinity ligand of HIV-1 TAR reveals the RNA structure recognized by P-TEFb. *Nucleic Acids Research*, 47(3), 1523-1531.
- Smirnov, A., Förstner, K., Holmqvist, E., Otto, A., Günster, R., & Becher, D. et al. (2016). Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proceedings Of The National Academy Of Sciences*, 113(41), 11591-11596.
- Smirnov, A., Schneider, C., Hör, J., & Vogel, J. (2017). Discovery of new RNA classes and global RNA-binding proteins. *Current Opinion In Microbiology*, 39, 152-160.
- Stelzer, A., Frank, A., Kratz, J., Swanson, M., Gonzalez-Hernandez, M., & Lee, J. et al. (2011). Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nature Chemical Biology*, 7(8), 553-559.
- Suryawanshi, H., Sabharwal, H., & Maiti, S. (2010). Thermodynamics of Peptide–RNA Recognition: The Binding of a Tat Peptide to TAR RNA. *The Journal Of Physical Chemistry B*, 114(34), 11155-11163.
- Tan, R., & Frankel, A. (1992). Circular dichroism studies suggest that TAR RNA changes conformation upon specific binding of arginine or guanidine. *Biochemistry*, 31(42), 10288-10294.
- Tao, J., & Frankel, A. (1992). Specific binding of arginine to TAR RNA. *Proceedings Of The National Academy Of Sciences*, 89(7), 2723-2726.
- Thomas, J., Ruggiero, A., Paxton, W., & Pollakis, G. (2020). Measuring the Success of HIV-1 Cure Strategies. *Frontiers In Cellular And Infection Microbiology*, 10.

Voss, S., Fischer, R., Jung, G., Wiesmüller, K., & Brock, R. (2006). A Fluorescence-Based Synthetic LPS Sensor. *Journal Of The American Chemical Society*, *129*(3), 554-561.

Wang, S., Huber, P., Cui, M., Czarnik, A., & Mei, H. (1998). Binding of Neomycin to the TAR Element of HIV-1 RNA Induces Dissociation of Tat Protein by an Allosteric Mechanism. *Biochemistry*, *37*(16), 5549-5557.

Wu, X., Lan, L., Wilson, D., Marquez, R., Tsao, W., & Gao, P. et al. (2015). Identification and Validation of Novel Small Molecule Disruptors of HuR-mRNA Interaction. *ACS Chemical Biology*, *10*(6), 1476-1484.

Zhang, X., Lee, S., Zhao, L., Xia, T., & Qin, P. (2010). Conformational distributions at the N-peptide/boxB RNA interface studied using site-directed spin labeling. *RNA*, *16*(12), 2474-2483.

Chapter 4

Conclusion and Future Outlook

1. General Conclusion

All the way back in 1998, the World Health Assembly adopted a resolution that urged its Member States to take action against the growing global antimicrobial resistance (AMR) (Leung, Weil, Raviglione & Nakatani, 2011). In 2005, the World Health Organization selected antimicrobial resistance as the theme for its World Health day, and even now continues to highlight AMR as one of the top 10 global public health threats facing humanity ("Antimicrobial resistance", 2021). A study done in the EU in 2015 estimated that in 2015 alone, there were 671 689 cases of infections with selected antibiotic-resistant bacteria that occurred in the EU and European Economic Area (EEA), and that the burden of resistant infections was similar to the combined burden of three major infectious diseases (influenza, tuberculosis, and HIV) (Cassini et al., 2019).

While tools to help detect growing AMR have been developed (for example, Rowe et al., 2015), what is sorely needed is new, innovative antimicrobials. With the pipeline for new drugs growing dry, new drug scaffolds must be developed to slow down the rate of growing resistance. Most of the newer drugs are modified versions of existing drugs, which make it easier for bacteria and viruses to gain resistance against (Kmietowicz, 2017). Hence, it would not be a stretch to say that the discovery of new drugs as well as new drug targets is vital for humanity's survival.

To that end, RNA is being raised up as a potential target to expand the druggable space (Ursu, Vézina-Dawod & Disney, 2019; Rizvi et al., 2020). With new chemical libraries of RNA-targeting small molecules being developed (Rzuczek, Southern & Disney, 2015), RNA-adaptable high-throughput screening methods will no doubt gain importance. We hope that the fluorescent indicator displacement (FID) assay would gain popularity as an orthogonal screening method to cellular-based assays due to its ease of use. However, the main limitation of the FID assay as a screening method is the need for a good indicator (Chapter 1.4).

This research thus hopes to offer more creative ideas to developing dsRNA-targeting probes. By introducing two new probe designs – SPOC and LUPI, we would like to see future research that expands on these designs or that take inspiration from them. Furthermore, from these experimental results, we hope to provide some insights into how the selectivity or binding affinity of these probes can be finetuned.

3. Future Outlook for LUPI

Light-Up Peptide Indicator (LUPI)

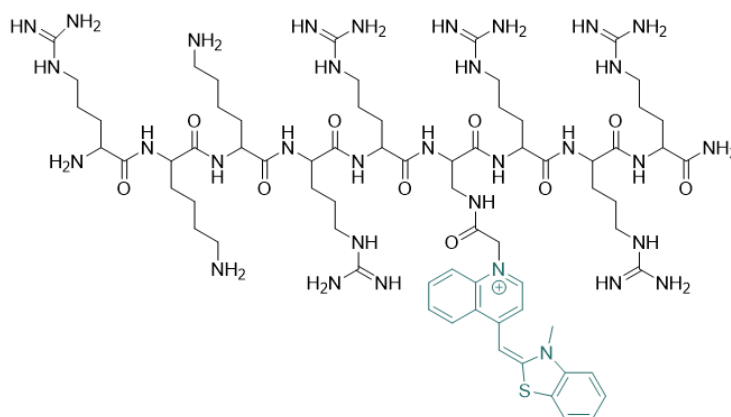


Figure 2. Chemical structure of LUPI.

Our second proposed probe design targeted ribonucleoprotein (RNP) complexes instead of just dsRNA itself. LUPI was designed using the wild-type RNA-binding protein's RNA binding domain's amino acid sequence, where one amino acid was replaced by a TO amino acid surrogate, the first of its kind to be reported. We chose the HIV-1 Tat protein – TAR RNA model as our proof-of-concept. One of the most noteworthy results was the incredibly strong binding affinity that LUPI had with the TAR RNA ($K_d = 1.0 \pm 0.6$ nM), allowing it to sieve out super strong binders in the mock FID assay. While there are other macrocyclic peptide mimics of Tat that have even lower K_d values (Shortridge et al., 2018), LUPI works well for our target application of FID assays.

The generality of the LUPI probe design can be shown by applying the LUPI concept to other RNP complexes. At present, we are thinking of applying it to other well-known RNPs, such as the HIV-1 Rev protein - Rev-response element region IIB (RRE IIB) RNA model (Kjems, Calnan, Frankel & Sharp, 1992), or the boxB RNA-λN protein complex (Rees, Weitzel, Yager, Das & von Hippel, 1996).

Furthermore, because LUPI binds so tightly to TAR RNA, it can be used for other applications as well. In Chapter 3, we attempted to use LUPI to image the nucleolus in MCF7 cells, but LUPI did not perform as well as other imaging dyes. LUPI's low LOD with TAR RNA (18 pM) could make it an attractive detection tool.

4. References

- Antimicrobial resistance. (2021). Retrieved 20 July 2022, from <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>
- Cassini, A., Högberg, L., Plachouras, D., Quattrocchi, A., Hoxha, A., & Simonsen, G. et al. (2019). Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *The Lancet Infectious Diseases*, *19*(1), 56-66.
- Das, B., Murata, A., & Nakatani, K. (2021). A small-molecule fluorescence probe ANP77 for sensing RNA internal loop of C, U and A/CC motifs and their binding molecules. *Nucleic Acids Research*, *49*(15), 8462-8470.
- Gupta, P., Zengeya, T., & Rozners, E. (2011). Triple helical recognition of pyrimidine inversions in polypurine tracts of RNA by nucleobase-modified PNA. *Chemical Communications*, *47*(39), 11125.
- Higuchi, K., Sato, Y., Togashi, N., Suzuki, M., Yoshino, Y., & Nishizawa, S. (2022). Bright and Light-Up Sensing of Benzo[c,d]indole-oxazolopyridine Cyanine Dye for RNA and Its Application to Highly Sensitive Imaging of Nucleolar RNA in Living Cells. *ACS Omega*, *7*(27), 23744-23748.
- Kjems, J., Calnan, B., Frankel, A., & Sharp, P. (1992). Specific binding of a basic peptide from HIV-1 Rev. *The EMBO Journal*, *11*(3), 1119-1129.
- Kmietowicz, Z. (2017). Few novel antibiotics in the pipeline, WHO warns. *BMJ*, j4339.
- Leung, E., Weil, D., Raviglione, M., & Nakatani, H. (2011). The WHO policy package to combat antimicrobial resistance. *Bulletin Of the World Health Organization*, *89*(5), 390-392.
- Ong, A., Toh, D., Patil, K., Meng, Z., Yuan, Z., & Krishna, M. et al. (2019). General Recognition of U-G, U-A, and C-G Pairs by Double-Stranded RNA-Binding PNAs Incorporated with an Artificial Nucleobase. *Biochemistry*, *58*(10), 1319-1331.
- Rees, W., Weitzel, S., Yager, T., Das, A., & von Hippel, P. (1996). Bacteriophage lambda N protein alone can induce transcription antitermination in vitro. *Proceedings Of The National Academy Of Sciences*, *93*(1), 342-346.
- Rizvi, N., Santa Maria, Jr., J., Nahvi, A., Klappenbach, J., Klein, D., & Curran, P. et al. (2020). Targeting RNA with Small Molecules: Identification of Selective, RNA-Binding Small Molecules Occupying Drug-Like Chemical Space. *SLAS Discovery*, *25*(4), 384-396.
- Rowe, W., Baker, K., Verner-Jeffreys, D., Baker-Austin, C., Ryan, J., Maskell, D., & Pearce, G. (2015). Search Engine for Antimicrobial Resistance: A Cloud Compatible Pipeline and Web Interface for Rapidly Detecting Antimicrobial Resistance Genes Directly from Sequence Data. *PLOS ONE*, *10*(7), e0133492.

Rzuczek, S., Southern, M., & Disney, M. (2015). Studying a Drug-like, RNA-Focused Small Molecule Library Identifies Compounds That Inhibit RNA Toxicity in Myotonic Dystrophy. *ACS Chemical Biology*, *10*(12), 2706-2715.

Shortridge, M., Wille, P., Jones, A., Davidson, A., Bogdanovic, J., & Arts, E. et al. (2018). An ultra-high affinity ligand of HIV-1 TAR reveals the RNA structure recognized by P-TEFb. *Nucleic Acids Research*, *47*(3), 1523-1531.

Ursu, A., Vézina-Dawod, S., & Disney, M. (2019). Methods to identify and optimize small molecules interacting with RNA (SMIRNAs). *Drug Discovery Today*, *24*(10), 2002-2016.

Yoshino, Y., Sato, Y., & Nishizawa, S. (2019). Deep-Red Light-up Signaling of Benzo[c,d]indole-Quinoline Monomethine Cyanine for Imaging of Nucleolar RNA in Living Cells and for Sequence-Selective RNA Analysis. *Analytical Chemistry*, *91*(22), 14254-14260.