NMR and SPR Fragment-Based Screening Can Produce Novel High Affinity Small Molecule Hits Against Structured RNAs

Brooke X.C. Kwai[1,2], Indu R. Chandrashekaran[1,2], Biswaranjan Mohanty[1,3], Menachem J. Gunzburg[1,4], Bradley C. Doak[1,4], Ashish Sethi[5], Shubhadra Pillay[6], David Lok[7], Sean J. Harrison[7], Pedro Serrano[7], Elisa Barile[6,#], and Martin J. Scanlon[1,2,4,*]

1. Medicinal Chemistry
   Monash Institute of Pharmaceutical Sciences
   Monash University
   Parkville, Victoria, 3052, Australia

2. ARC Centre for Fragment-Based Design
   Monash University
   Parkville, Victoria, 3052, Australia

3. Sydney Analytical Core Research Facility
   The University of Sydney
   Sydney, New South Wales, 2006, Australia

4. Monash Fragment Platform
   Monash Institute of Pharmaceutical Sciences
   Monash University
   Parkville, Victoria, 3052, Australia

5. Australian Synchrotron
   Australian Nuclear Science and Technology Organisation
   Clayton, Victoria, 3168, Australia

6. Takeda California, Inc.
   9625 Towne Centre Drive, San Diego, 92121, USA

7. Takeda Development Centre Americas, Inc
   40 Landsdowne St., Cambridge, MA, 02478, USA

[#]E-mail: elisa.barile@takeda.com
*E-mail: martin.scanlon@monash.edu

Abstract

Non-coding RNAs account for up to 98 % of the human transcriptome.[1] It has become increasingly clear that non-coding RNAs play diverse and critical roles in many important cellular functions.[2, 3] Although modulation of non-coding RNAs using small molecules is a promising therapeutic strategy, there are relatively few well-characterised RNA-ligand structures. Therefore, the structure-interaction relationships of RNA-targeting small molecules remain underexplored.[4] Here we present a fragment-based screening approach using biophysical assays to identify and evaluate fragments that bind to the theophylline RNA aptamer, which we use as a model system. We were able to identify high affinity fragment hits and generate models of RNA-ligand complexes using a combination of biophysical data and computational docking. Together, these provided insights into the RNA-fragment interactions

that underpin binding. This approach demonstrates the feasibility of identifying high-affinity RNA-targeting small molecules with limited structural information.

Introduction

Non-coding RNAs (ncRNAs) are a largely unexplored target space for drug discovery. Although ncRNAs make up the majority of the human transcriptome and significantly outnumber protein-coding genes,[1] almost all current therapeutic small molecules on the market are designed to target disease-associated proteins.[4, 5] ncRNAs are involved in a variety of complex roles in biological processes. They can regulate metabolite pathways and gene expression. Growing evidence indicates that ncRNAs play a role in various diseases, including cancers, heart disease and neurodegeneration. Risdiplam is an RNA-targeting drug that has been clinically approved for the treatment of spinal muscular atrophy. Risdiplam modifies splicing of the survival of motor-neurone (SMN) 2 pre-mRNA, thereby increasing the expression of the SMN protein.[6] This demonstrates the potential of targeting RNA for drug development.[7, 8]

A range of different methods has been described for identifying promising small molecule leads that bind to RNA.[9] However, deciphering the structural basis for selective binding to RNA remains difficult and only a handful of specific RNA-binding small molecules have been developed.[4] The dynamic nature of RNAs, their polyanionic backbone and lack of deep hydrophobic pockets for binding small organic compounds contribute to a tendency for RNA ligands to be weak, non-specific, positively charged compounds.[10] Moreover, there are relatively few well-characterised atomic resolution structures of RNA. This makes target- and structure-based design approaches, which are used routinely against protein targets, more challenging for RNA. More robust methods for generating specific RNA-binding small molecules would be of tremendous benefit given the growing interest in RNA therapeutics.

In this work, we have used the theophylline aptamer as a model system to evaluate fragment-based screening (FBS) approaches to RNA-ligand identification. FBS is a powerful approach for hit-finding as it allows efficient sampling of the chemical space and diverse chemotype screening.[11] FBS has been shown to be effective in identifying hits for 'difficult-to-target' proteins.[12] Therefore, it is potentially well-suited to identifying ligands for RNAs, which are also challenging targets. Herein, we report a fragment-based screening method using NMR and SPR that was used to identify several series of ligands that bound to the theophylline aptamer, some with higher affinity than theophylline. We generated computational models of the aptamer in complex with ligands from our best series, which suggested that chemically diverse scaffolds bind through a similar interaction network.

Results and Discussion

To identify fragments that bound to the aptamer, we used NMR spectroscopy to monitor the chemical shifts of the imino protons of the RNA. These imino protons form G-C and A-U Watson-Crick base pairs and have chemical shift resonances that are observed between 10 to 15 ppm in their ^{1}H NMR spectra. These downfield resonances are excellent probes for RNA folding and can also be used in ligand binding studies. They are usually well resolved and sufficiently distant in chemical shift from small molecule resonances that there is little peak overlap. They are also extremely sensitive to fragment binding, which means that binding can be identified based on chemical shift perturbations (CSP) or peak broadening. We took advantage of these unique properties and used RNA-detected 1D ^{1}H NMR spectroscopy as a primary screening tool to identify fragments that bound to the theophylline aptamer. For NMR studies we used the △TCT8-4 construct of the theophylline-binding aptamer (Supplementary Figure 1), which in its apo form shows good stability in solution, as indicated by its NMR

spectrum that remains unchanged over a period of seven days (data not shown). The apo △TCT8-4 aptamer spectrum was used as a reference for the NMR screen. A 1D ^{1}H NMR spectrum was recorded for the △TCT8-4 aptamer in the presence of either theophylline or caffeine to confirm that the aptamer was able to bind theophylline but not caffeine. The spectrum showed clear CSPs upon the addition of theophylline, consistent with those reported previously, indicating that theophylline bound under the conditions used for screening.[13] In contrast, no CSPs were observed with the addition of caffeine. These spectra validate the screening method and serve as negative and positive controls (Figure 1A). We used a library of 1975 commercial fragments, selected for diversity, that were soluble and free from aggregation in aqueous buffer. The library was divided into 320 cocktails of six fragments and 11 cocktails of five fragments for screening. 1D ^{1}H NMR spectra of the △TCT8-4 aptamer in the presence of fragment mixtures were inspected manually for hit identification. A mixture was considered to contain a hit fragment if it induced a CSP of an RNA imino proton resonance of greater than 0.05 ppm or if it resulted in a change in the number of clearly observable imino proton resonances in the spectrum. Using these criteria, the primary screen identified 25 hit mixtures.

The hits were deconvoluted by recording 1D ^{1}H NMR spectra of the aptamer in the presence of each of the single fragments from the 25 hit mixtures (Figure 1B). A total of 28 fragments were validated as hits from the primary screen (Supplementary Figure 2). This corresponds to an initial hit rate of 1.4 %. The screening cascade is summarised as a flowchart in Figure 1C. Based on the low hit rate, we also explored the possibility of using larger cocktails to accelerate screening. We made 12-fragment mixtures by combining two six-fragment cocktails and found that it was possible to reliably identify and deconvolute all hits with this format (Supplementary Figure 3), even in cases where there was more than one hit in the mixture. Through doubling the number of fragments in a primary screen mixture the total NMR time required for a primary screen would be reduced by half. While twice as many experiments are required for deconvolution in this format, the low hit rate dictates that the overall process is accelerated. Hence, the optimal number of fragments contained in a screening mixture is dependent on the hit rate with lower hit rates benefiting from larger mixtures.

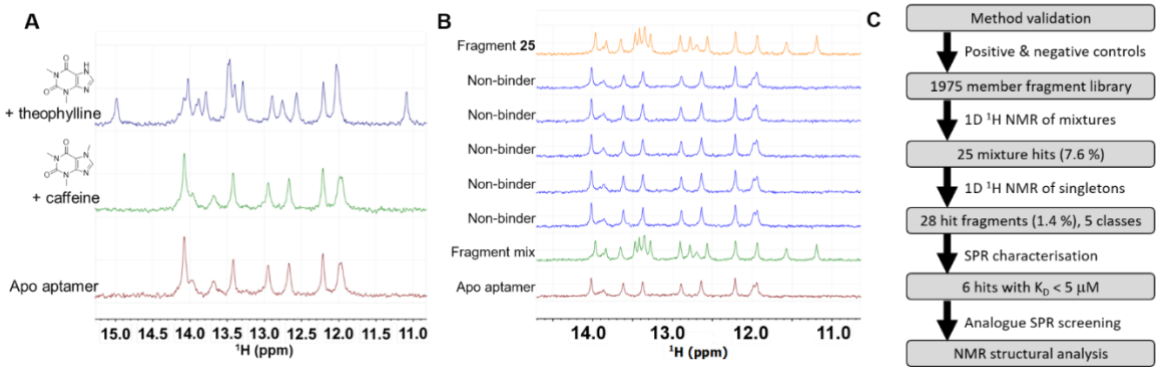

***Figure 1**. Overview of the fragment-based screen. (A) 1D ^{1}H NMR spectra of 36 µM ΔTCT8-4 aptamer imino region in the absence of ligand (apo, red) and in the presence of 360 µM theophylline (blue) and 360 µM caffeine (green). (B) Hit deconvolution using NMR. 1D ^{1}H NMR imino region of the ΔTCT8-4 aptamer in the absence of fragments (red), in the presence of a mixture of 6 fragments (each at 312.5 µM) (green), and in the presence of each individual fragment (1000 µM) of the mixture. Non-binding fragments are in blue, and the hit fragment is in orange. (C) Flowchart summarising the fragment screening cascade.*

Dose-response analysis was carried out by recording 1D ^{1}H NMR spectra of the aptamer in the presence of each hit fragment at concentrations of 50 µM, 200 µM and 625 µM,

respectively. All but three of the fragments were observed to be in slow exchange on the NMR timescale, with the remaining three fragments showing intermediate exchange with the aptamer. Therefore, to estimate the binding affinity of the hit fragments, we employed surface plasmon resonance (SPR) as a secondary screening tool (Supplementary Table 1). Biotinylated theophylline aptamer was immobilised onto a streptavidin-coated SA sensor chip (Cytiva). Theophylline was used as a positive control in the assay, providing a measured K_D = 420 nM. This is consistent with the previous literature, where the reported K_D = 320 nM.[13] Four of the 28 aptamer-binding fragments were found to have a sub-micromolar binding affinity measured by SPR. They displayed saturation and showed 1:1 binding in the tested concentration range. Both kinetic and thermodynamic analyses for these four fragments gave a consistent estimate of the binding affinity (Figure 2). All four fragments have a similar bicyclic and pyrimidinone-like motif, bearing some resemblance to theophylline. Fragment **27** was identified to have a higher affinity than theophylline (SPR K_D = 90 nM).

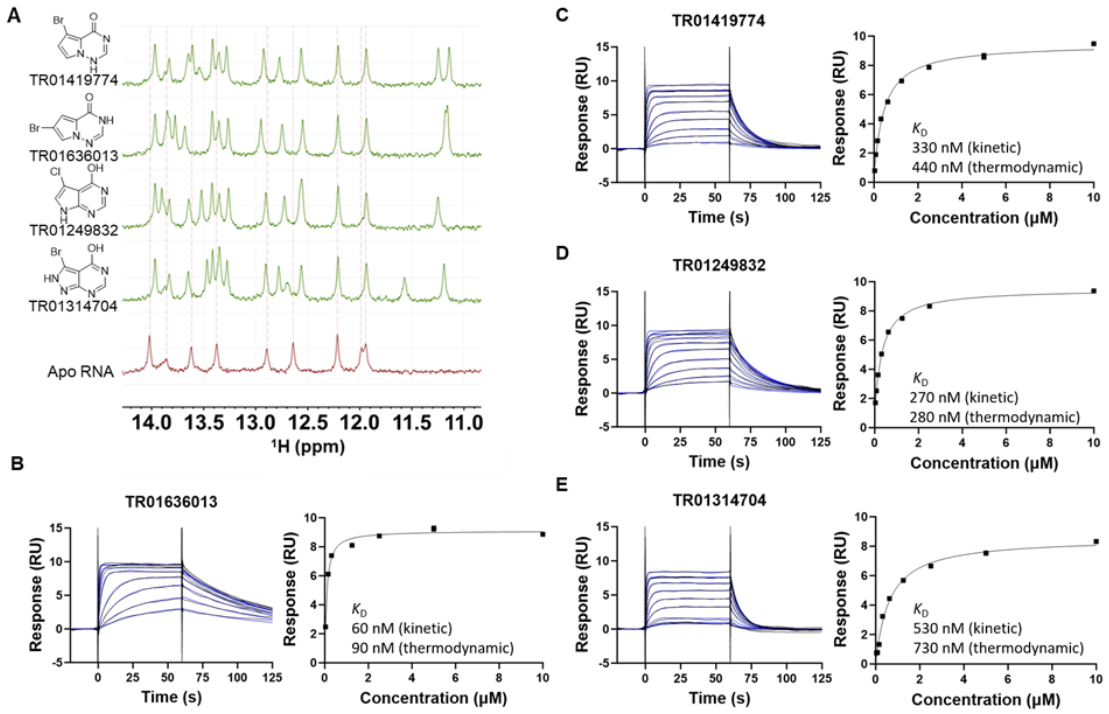

***Figure 2**. 1D ^{1}H NMR and SPR profile of the top four primary screen singleton hits. (A) Comparison between the imino region of the 1D ^{1}H NMR spectra of the apo ΔTCT8-4 aptamer (36 μM) in red, and in the presence of each individual fragment (1000 μM) in green. Dotted lines indicate the position of the apo ΔTCT8-4 aptamer chemical shifts. (B-E) SPR sensorgrams of ligand binding (0-100 μM) to the ΔTCT8-4 aptamer with corrected response in black lines and fitted model in blue lines fitted to 1:1 kinetics binding model and with thermodynamic analysis fitted to steady-state affinity model.*

The aptamer's tolerance to the different patterns of substitution on the bicyclic core was surprising, considering the specificity for theophylline over caffeine, where addition of a single methyl group results in >10,000-fold loss of affinity.[14] To investigate the structure-binding relationships further, we obtained 14 structural analogues of fragment **27** to identify permitted substitution patterns (Figure 3). Dose-response 1D ^{1}H NMR (Supplementary Figure 4) and SPR data were generated for these compounds. 1D ^{1}H NMR identified 12 fragments that were in slow exchange on the NMR time scale. Fragment **29** is a methyl analogue of its parent

fragment **27**. It has the highest affinity (SPR K_D = 370 nM) among the 14 analogues (Supplementary Figure 5), albeit having slightly lower affinity than fragment **27**.

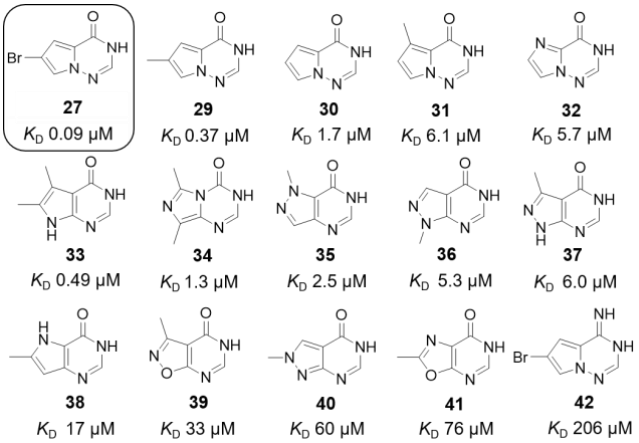

***Figure 2**. Structures of the 14 analogues of fragment 27 (shown in box) and binding affinities as measured by thermodynamic SPR.*

To gain an insight into the binding mode of the fragments, we recorded 2D ^{1}H-^{1}H NOESY NMR spectra of the unlabelled △TCT8-4 aptamer in complex with both theophylline and fragment **27**, which was the highest affinity compound identified from the screen. Initially, proton resonance assignments for imino protons in the △TCT8-4 aptamer were made by comparison with those published for the △-33 spectrum in the presence of theophylline.[15] We recorded 2D ^{1}H-^{1}H NOESY for theophylline-bound △TCT8-4. Most of the imino-imino NOE cross peaks for the paired guanines and uracils could be successfully assigned for the △TCT8-4 conserved region, by comparison with the literature values for the △-33 spectrum (Supplementary Figure 1B). Interactions between the theophylline methyl groups H1 and H3 as well as H7 and the imino protons of the conserved residues, U7, U29, U30 and G32 at the binding site in △TCT8-4 were unchanged from those reported for the complex of theophylline-bound to △-33 (Figure 4). This suggests that the △TCT8-4 aptamer adopts a similar structure within the theophylline binding pocket as △-33, despite the slight differences in the sequence; and theophylline is likely to bind in a similar orientation across the two aptamers.

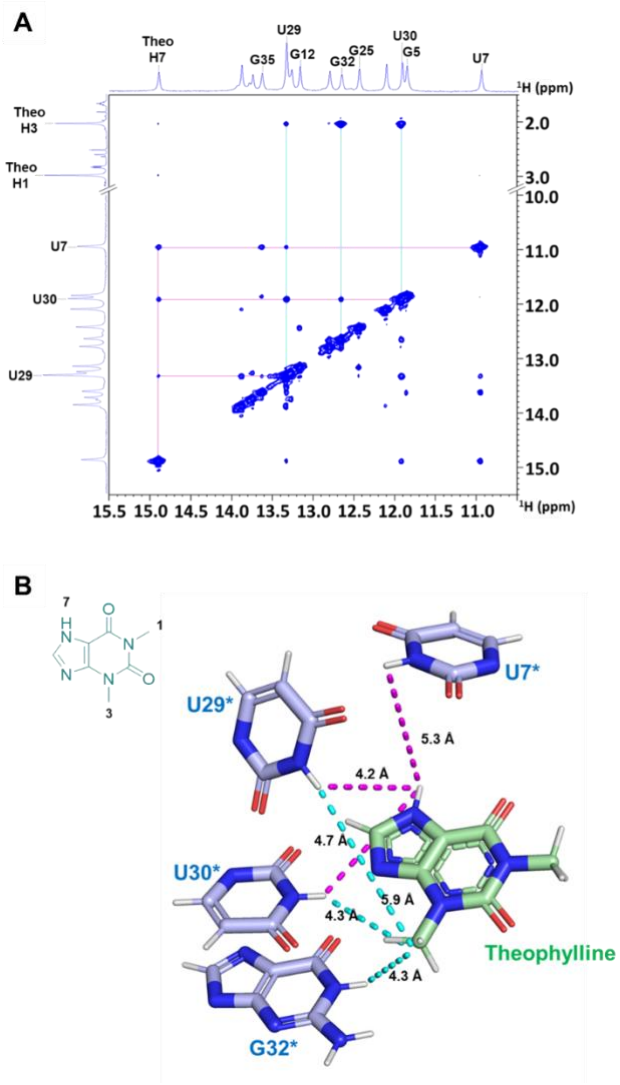

***Figure 3**. Binding mode of theophylline derived from NMR data. (A) 2D ^{1}H-^{1}H NOESY spectrum of the ΔTCT8-4 aptamer (375 μM) bound to theophylline (500 μM). Assignments for the imino proton resonances of ΔTCT8-4 aptamer and key resonances of theophylline are labelled on the 1D projections of the spectra. Intermolecular NOE cross peaks to H7 of theophylline are shown in magenta, and to H3 of theophylline are shown in cyan. (B) the binding orientation of theophylline bound to the Δ-33 aptamer (PDB 1EHT). The intermolecular NOE correlations observed from H7 of theophylline are shown in magenta, and the intermolecular NOE correlations observed from H3 of theophylline are shown in cyan. * Bases are labelled according to the sequence of ΔTCT8-4.*

Next, we recorded small angle X-ray scattering data (SAXS) for △TCT8-4 in the presence of theophylline, fragment **27** or fragment **31** as representatives of the compound series (Supplementary Figure 6 and Supplementary Table 2). Analysis of the SAXS data revealed that there was no major change in the conformation of the aptamer in the presence of theophylline or the two ligands identified in this work. Therefore, we were confident we could apply a similar analysis of the NOESY data for the complex of △TCT8-4 in the presence of the 4- pyrimidinone-like fragments.

The highest affinity analogue fragment **27** contains only aromatic protons, which are present in a congested region of the NMR spectrum. This makes interpretation of intermolecular cross peaks in the NOESY spectrum challenging. Nonetheless, a similar pattern of intramolecular

NOEs between the well-resolved imino protons of the △TCT8-4 aptamer theophylline binding site was observed in the presence of both theophylline and fragment **27** (Supplementary Figure 7). NOESY data were also acquired for the aptamer complex with fragment **29** (K_D = 370 nM). It contains a methyl group at C-6 that resonates in a well-resolved region in the NMR spectrum which aids the interpretation of the fragment orientation by providing unambiguous intermolecular NOEs for the complex (Figure 5 & Supplementary Figure 8).

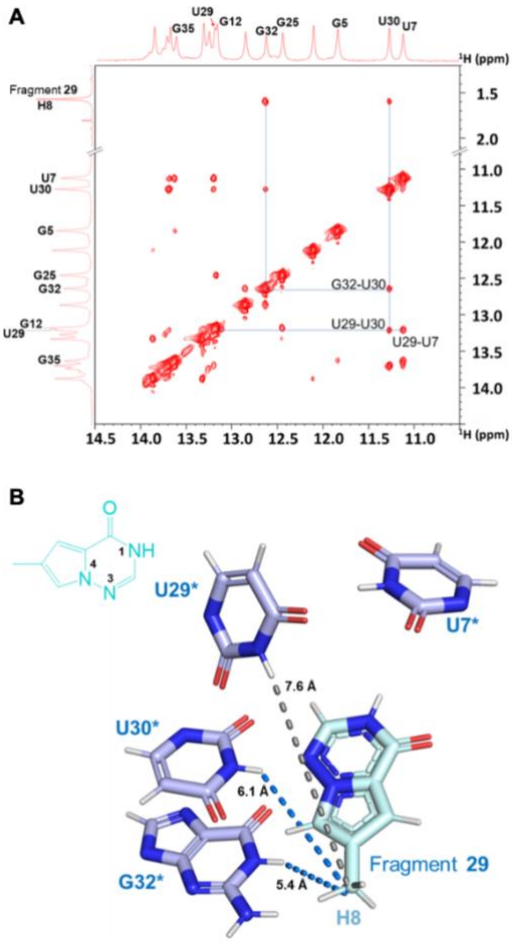

Figure 4*. Binding mode of fragment 29 derived from NMR data. (A) 2D ^{1}H-^{1}H NOESY spectrum of the fragment 29 bound ΔTCT8-4 aptamer. Assignments for the imino proton resonances of the RNA and the methyl of fragment 29 are labelled on the 1D projections of the spectra. Intermolecular NOE cross peaks are labelled. (B) fragment 29 docked to the Δ-33 aptamer (PDB 1EHT). The observed intermolecular NOE correlations to H8 of the fragment are shown as blue dotted lines. The distance between U29 to H8 of the fragment that is shown as a grey dotted line. * Bases are labelled according to the sequence of ΔTCT8-4.*

The imino-imino and imino-methyl regions of the NOESY spectra for U7, U29, U30 and G32 of △TCT8-4 in complex with fragment **29** are presented in Figure 5A. Strong NOEs are observed between the methyl group H8 of fragment **29** and the imino protons of U30 and G32. In contrast to the data observed for the theophylline N-3 methyl, there is no NOE observed between the methyl protons of fragment 29 and the imino proton of U29. The assignment of the imino protons of U7, U29, U30 and G32 in complex with fragment 29 is supported by the network of imino-imino NOEs that are similar to those observed in the presence of theophylline (Figures 4A & 5A). Molecular docking of fragment **29** to the binding pocket of △-33 (PDB

1ETH) was employed to rationalise the observed NOEs. Together, the docking results and the NOE correlations observed suggest that fragment **29** is binding with its methyl group pointing away from U29 (△TCT8-4 numbering, Figure 5B). In this orientation, fragment **29** maintained a similar network of H-bonds to those observed in the structure of theophylline with the equivalent bases (Figure 6). The docking results suggest the NH protons of C27 and U30 act as H-bond donors in interactions with the pyrimidinone core nitrogen in fragment **29**, while the NH2 of C27 acts as a H-bond donor to the oxygen of fragment **29**. This also positions fragment **29** in a suitable orientation to maintain the π- π stacking interactions observed between the aptamer and theophylline.

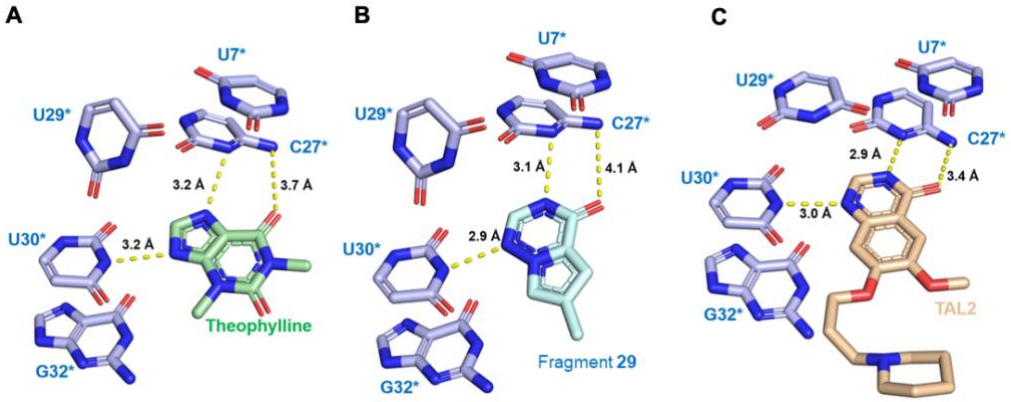

***Figure 5**. Different chemical scaffolds maintain a conserved mode of interaction with theophylline aptamers. (A) NMR structure of theophylline bound to the Δ-33 aptamer (PDB 1EHT) with hydrogen bonds between theophylline and the aptamer bases shown as yellow dotted lines. (B) The docked structure of fragment 29 bound to the Δ-33 aptamer (PDB 1EHT) showing hydrogen bonds between the fragment and the aptamer bases shown as yellow dotted lines. (C) Crystal structure of TAL2 bound to a theophylline aptamer variant (PDB 8D2B) with hydrogen bonds between TAL2 and the aptamer bases shown as yellow dotted lines. * All bases are labelled according to the sequence of ΔTCT8-4.*

Fragment **31** is another methyl-containing analogue of fragment **27**, which binds with ~7-fold lower affinity (K_D = 610 nM) than fragment **29**. NMR spectra were acquired for fragment **31** bound to △TCT8-4 (Supplementary Figure 9). Assignments for the imino protons in the complex were made by comparison of the patterns of NOEs in imino-aromatic and imino-imino regions of the NOESY spectrum. In the case of fragment **31**, there were no NOEs observed between its methyl protons and any of the imino protons of the aptamer (Supplementary Figure 9B). A low-energy docking model of the complex suggested that fragment **31** was bound in a similar orientation as fragment 29, with the 6-membered ring forming three hydrogen bonding with U30 and C27 (Supplementary Figure 9E). In this orientation, the methyl protons were more than 6.7 Å away from the imino proton of the closest base G32, which supports the lack of NOEs observed for this complex. Therefore, the NOESY data were consistent with the docking models for each of these two analogues.

Despite the lack of interpretable NOEs in the spectrum of fragment **27** bound to △TCT8-4, the patterns of NOEs and the CSP observed in the NMR spectra are similar in the presence of either theophylline or any of the three fragments (Supplementary Figures 7, 8 and 9). These NMR data are consistent with the pyrimidinone core of fragment **27** adopting a similar orientation to that of **29** and **31**.

The mode of binding derived from the docking and NMR analysis for the pyrimidinone-like fragments indicates that the pyrimidinone core maintains similar interactions with the RNA as

both theophylline and a quinazolinone-containing ligand (TAL2) that was reported by Novartis (Figure 6).[10] The quinazolinone binding mode was identified from the X-ray crystal structure of TAL2 bound with a modified △-33 aptamer (PDB 8D2B). Together this suggests that a similar interaction network can support high affinity binding of multiple different scaffolds to the theophylline-binding aptamer.

In this work, we have demonstrated the feasibility of using a general, non-RNA specific fragment library to conduct a fragment-based screen against the theophylline aptamer. Our FBS method successfully identified a chemical scaffold that bound with high affinity (K_D = 90 nM) to an aptamer that is considered to be highly selective for theophylline. We could extract adequate structural information using unlabelled RNA through 2D NOESY and SAXS, in conjunction with docking data to establish a mode of binding. The overall hit rate of 1.4 % is relatively low, which is consistent with the expectation that RNAs represent challenging targets.[16] However, our ability to find ligands that bind to the aptamer with higher affinity than theophylline suggests that this workflow is potentially applicable towards a diverse range of RNA targets. It can be translated into an efficient tool for generating RNA-binding small molecule as chemical probes and lead compounds for drug development.

Acknowledgements

This research was conducted by the ARC Industrial Transformation Training Centre for Fragment-Based Design (Project ID IC180100021) and funded by the Australian Government.

References

(1) Dunham, I.; Kundaje, A.; Aldred, S. F.; Collins, P. J.; Davis, C. A.; Doyle, F.; Epstein, C. B.; Frietze, S.; Harrow, J.; Kaul, R.; et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489* (7414), 57-74. DOI: 10.1038/nature11247

(2) Zheleznyakova, G. Y.; Piket, E.; Needhamsen, M.; Hagemann-Jensen, M.; Ekman, D.; Han, Y.; James, T.; Khademi, M.; Al Nimer, F.; Scicluna, P.; et al. Small noncoding rna profiling across cellular and biofluid compartments and their implications for multiple sclerosis immunopathology. *Proc. Natl. Acad. Sci.* **2021**, *118* (17), e2011574118. DOI: doi:10.1073/pnas.2011574118

(3) Statello, L.; Guo, C.-J.; Chen, L.-L.; Huarte, M. Gene regulation by long non-coding rnas and its biological functions. *Nat. Rev. Mol. Cell Biol.* **2021**, *22* (2), 96-118. DOI: 10.1038/s41580-020-00315-9

(4) Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principles for targeting rna with drug-like small molecules. *Nat. Rev. Drug Discovery* **2018**, *17* (8), 547-558. DOI: 10.1038/nrd.2018.93

(5) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discovery* **2017**, *16* (1), 19-34. DOI: 10.1038/nrd.2016.230

(6) Ratni, H.; Ebeling, M.; Baird, J.; Bendels, S.; Bylund, J.; Chen, K. S.; Denk, N.; Feng, Z.; Green, L.; Guerard, M.; et al. Discovery of risdiplam, a selective survival of motor neuron-2 (smn2) gene splicing modifier for the treatment of spinal muscular atrophy (sma). *J. Med. Chem.* **2018**, *61* (15), 6501-6517. DOI: 10.1021/acs.jmedchem.8b00741

(7) Zhang, Y.; Mao, Q.; Xia, Q.; Cheng, J.; Huang, Z.; Li, Y.; Chen, P.; Yang, J.; Fan, X.; Liang, Y.; et al. Noncoding rnas link metabolic reprogramming to immune microenvironment in cancers. *J. Hematol. Oncol.* **2021**, *14* (1), 169. DOI: 10.1186/s13045-021-01179-y

(8) Zhu, L.; Li, N.; Sun, L.; Zheng, D.; Shao, G. Non-coding rnas: The key detectors and regulators in cardiovascular disease. *Genomics* **2021**, *113* (1, Part 2), 1233-1246. DOI: https://doi.org/10.1016/j.ygeno.2020.10.024

(9) Childs-Disney, J. L.; Yang, X.; Gibaut, Q. M. R.; Tong, Y.; Batey, R. T.; Disney, M. D. Targeting rna structures with small molecules. *Nat. Rev. Drug Discovery* **2022**, *21* (10), 736-762. DOI: 10.1038/s41573-022-00521-4

(10) Menichelli, E.; Lam, B. J.; Wang, Y.; Wang, V. S.; Shaffer, J.; Tjhung, K. F.; Bursulaya, B.; Nguyen, T. N.; Vo, T.; Alper, P. B.; et al. Discovery of small molecules that target a tertiary-structured rna. *Proc. Natl. Acad. Sci.* **2022**, *119* (48), e2213117119. DOI: doi:10.1073/pnas.2213117119

(11) Murray, C. W.; Rees, D. C. The rise of fragment-based drug discovery. *Nat. Chem.* **2009**, *1* (3), 187-192. DOI: 10.1038/nchem.217

(12) Doak, B. C.; Norton, R. S.; Scanlon, M. J. The ways and means of fragment-based drug design. *Pharmacol. Ther.* **2016**, *167*, 28-37. DOI: https://doi.org/10.1016/j.pharmthera.2016.07.003

(13) Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. High-resolution molecular discrimination by rna. *Science* **1994**, *263* (5152), 1425-1429. DOI: doi:10.1126/science.7510417

(14) Zimmermann, G. R.; Wick, C. L.; Shields, T. P.; Jenison, R. D.; Pardi, A. Molecular interactions and metal binding in the theophylline-binding core of an rna aptamer. *RNA* **2000**, *6* (5), 659-667. DOI: 10.1017/s1355838200000169

(15) Zimmermann, G. R.; Jenison, R. D.; Wick, C. L.; Simorre, J.-P.; Pardi, A. Interlocking structural motifs mediate molecular discrimination by a theophylline-binding rna. *Nat. Struct. Biol.* **1997**, *4* (8), 644-649. DOI: 10.1038/nsb0897-644

(16) Edfeldt, F. N.; Folmer, R. H.; Breeze, A. L. Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov. Today* **2011**, *16* (7-8), 284-287. DOI: 10.1016/j.drudis.2011.02.002