Published in final edited form as:

Angew Chem Int Ed Engl. 2017 October 16; 56(43): 13498–13502. doi:10.1002/anie.201707641.

Discovery of Key Physicochemical, Structural, and Spatial Properties of RNA-Targeted Bioactive Ligands

Brittany S. Morgan^[a], Jordan E. Forte^[a], Rebecca N. Culver^[a], Dr. Yuqi Zhang^[b], and Prof. Dr. Amanda E. Hargrove^[a]

^[a]Department of Chemistry, Duke University, Durham, NC 27708-0346 (USA)

^[b]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, 92037 (USA)

Abstract

While myriad non-coding RNAs are known to be essential in cellular processes and misregulated in diseases, the development of RNA-targeted small molecule probes has met with limited success. To elucidate guiding principles for selective small molecule:RNA recognition, we analyzed cheminformatic and shape-based descriptors for 104 RNA-targeted ligands with demonstrated biological activity (RNA-targeted Bioactive ligand Database, R-BIND). We then compared R-BIND to both FDA-approved small molecule drugs and RNA ligands without reported bioactivity. Several striking trends emerged for bioactive RNA ligands, including: i) compliance to medicinal chemistry rules; ii) distinctive structural features; and iii) enrichment in “rod-like” over other shapes. This work provides unique insights that directly facilitate the selection and synthesis of RNA-targeted libraries with the goal of efficiently identifying selective small molecule ligands for therapeutically relevant RNAs.

Table of Contents

A space for RNA: The curation and analysis of the RNA-targeted Bioactive ligand Database (R-BIND) uncovered distinguishing 2D and 3D molecular properties that support the existence of an “RNA-privileged chemical space.” These guiding principles can facilitate the future design of selective ligands targeted toward therapeutically relevant RNAs.

Keywords

chemical probes; cheminformatics; molecular recognition; principal moments of inertia; RNA recognition

The identification of regulatory RNAs, including those that may have therapeutic relevance, has exploded in recent years.^[1] Indeed, functional non-coding RNAs have been implicated in diseases ranging from bacterial^[2] and viral infections^[3] to metastatic cancer.^[4] At the same time, methods for the molecular characterization of these RNAs are more difficult and/or lacking compared to those for proteins.^[5] Chemical probes, for example, have greatly progressed the study of many protein classes and related diseases^[6] but have been challenging to develop for non-ribosomal RNAs. Slow progress in this field has been attributed, in part, to a lack of guiding principles for achieving specific small molecule:RNA interactions.^[7]

Previous attempts to quantify distinct properties of RNA ligands have been notable but limited. For example, a cheminformatic analysis by Aboul-ela et al. in 2009 identified small differences between RNA-binding ligands and other libraries, including FDA-approved drugs.^[7a, 8] In studies targeting r(CUG)-repeat sequences in 2015, Disney et al. tested a drug-like, RNA-biased library of bis-benzimidazoles and similar core structures and found that bioactive molecules had statistically significant differences in topological polar surface area (tPSA) as well as hydrogen bond acceptors (HBA) and donors (HBD).^[9] Inspired by this previous work and the recent surge in discoveries of biologically active RNA-targeting ligands,^[10] we proposed to more broadly analyze the key guiding principles for specific RNA-targeting. In this work, we report the discovery of distinct physicochemical, structural, and spatial properties of RNA-targeted bioactive ligands, which are expected to facilitate the rapid discovery of RNA-targeted chemical probes and subsequent evaluation of the therapeutic potential of non-ribosomal RNAs.

First, we gathered reports from reviews^[10-11] and recent primary literature that described organic small molecule probes for non-ribosomal RNAs with evidence of both *in vitro* binding and target engagement in cell culture or animal models (SI-1). To focus on small organic molecules, peptides and oligonucleotides were omitted.^[12] Aminoglycosides were excluded due to known promiscuous RNA-binding behavior.^[10c, 11b, 13] Molecules satisfying the selection criteria and highlighted in the conclusion of each report were included in the RNA-targeted Bioactive ligand Database (R-BIND) (available as RBIND.xls). While compiling R-BIND, we identified two strategies utilized to target RNA in biological systems: monovalent small molecules (SM) and multivalent ligands (MV) (Figure 1). We generally classified MV ligands as those with the presence of an alkyl, aryl, or peptide-like linker between two or more binding moieties and a molecular weight greater than 500 amu (SI-1a).

To analyze the physicochemical properties of these and other ligands, we adapted a list of 20 cheminformatic parameters from Tan and co-workers (SI Table 2-1).^[16] Before analysis, all molecules were unbiasedly adjusted to their major protomeric and tautomeric structures (pH = 7.4) using ChemAxon (SI 2). In our first analysis, we compared R-BIND (SM) and R-

To identify distinctive properties of bioactive RNA ligands, we compared R-BIND to two additional libraries. First, we compiled all RNA-binding small molecules from the Nucleic Acid Ligand Database (NALDB)^[17] and filtered the library analogously to R-BIND (SI-1). Small molecules present in R-BIND were removed from NALDB (SM) and (MV) to represent RNA-binding ligands without reported bioactivity (SI-1). In addition, the FDA-approved chemical entities list was downloaded from DrugBank^[18] and filtered to accurately compare organic small molecules (SI-1). The FDA library represented both “drug-like” space and “protein-targeted” space as ~90% of FDA-approved drugs target proteins.^[19]

As a preliminary comparison, principal component analysis (PCA) was used to reduce the dimensionality of the data and visualize the relative relationships between the libraries.^[16] PCA defines orthogonal axes, i.e. principal components (PCs), that represent the maximum variance in the data in an unbiased manner, prior to defining which data points are in which groups. In this analysis, the first three PCs described 67% of the variance within the data (SI Table 4-1). Generally, PC 1 was heavily influenced by molecular weight and related parameters, PC 2 by solubility and polarity-related parameters, and PC 3 by differences in aromatic rings and sp³-hybridized carbons (SI Table 4-2). Most small molecules clustered near the origin while the larger and multivalent molecules extended along the positive PC 1 axis (Figure 2, SI Figure 4-2). In comparison, the NALDB (SM) and R-BIND (SM) occupied a similar region, though R-BIND (SM) occupied a smaller subset of FDA chemical space along PC 1 and PC 3. To assist in determining the overlap between the libraries, we performed nearest neighbor clustering analysis (SI Table 5-1 and 5-2). Using 95% of the variance (PC 1-10), notable overlap was observed between the SM libraries with 13% of R-BIND (SM) ligands and 19% of NALDB (SM) ligands falling within the other library's cluster. Further, 31% of R-BIND (SM) and 9% of the FDA library overlapped with the other

library's cluster, suggesting that bioactive RNA-binding ligands can possess drug-like properties.

To determine the properties responsible for these similarities and differences, we again used a Mann-Whitney U test to compare the parameter distributions. For the comparison of R-BIND (SM) and FDA, we included only FDA molecules within the molecular weight range of R-BIND (SM) to account for the presence of larger and potentially multivalent molecules within the FDA library (SI Figure 3-2, SI Table 3-3). Interestingly, 4/7 medicinal chemistry properties (molecular weight (MW), HBA, RotB, tPSA) did not show significant differences ($P > 0.05$) while statistically significant increases were observed for water solubility (LogP, LogD) and HBD in R-BIND (SM). Among the molecular recognition properties, highly significant differences ($P < 0.001$) were observed with increased TC and a higher yet narrower range of accessible surface area (ASA). Further, there were statistically significant differences in the molecular complexity of the molecules, where R-BIND (SM) contained fewer sp^3 -hybridized carbons (Fsp³) and stereocenters (nStereo) compared to the FDA library. We cannot determine, however, the extent to which these differences can be attributed to reduced complexity in synthetic and commercial screening libraries, from which many R-BIND (SM) ligands were discovered.^[20]

Strikingly, the differences in distribution of 8/9 structural parameters were found to be highly statistically significant between R-BIND (SM) and FDA libraries (Figure 3, SI Figure 3-2, SI Table 3-3). These differences included atom counts where concurrently higher N count and lower O count were observed for R-BIND (SM). While the differences in ring count were modest, the ring types were different, and R-BIND (SM) had more aromatic rings (ArRings) and heteroatom-containing rings (HetRings). Collectively, these observations signify that monovalent biologically active RNA-binding ligands can possess “drug-like” properties but that these properties have thus far been achieved through distinct chemical architectures relative to FDA-approved drugs.

To compare differences between RNA-binding ligands with and without reported bioactivity, the Mann Whitney U test was repeated to compare the distributions of R-BIND (SM) and NALDB (SM) (SI Table 3-5). All medicinal chemistry and molecular recognition properties were found to not have statistically significant differences with the exception of relative polar surface area (relPSA). In contrast, all of the structural parameter distribution differences were found to be statistically significant except SysRR and Fsp³ (Figure 3, SI Figure 4). Repeating the analysis using the same molecular weight restriction as the FDA library (140-590 amu) yielded similar statistical results (SI Table 3-6). These comparisons further our hypothesis that the structural composition of small molecules is crucial for the specific recognition of RNA in biological systems.

As shape is known to be a critical component of RNA:protein recognition,^[21] we also analyzed the spatial properties of R-BIND (SM), FDA, and NALDB (SM) using calculated principal moments of inertia (PMI) (SI-6).^[22] The normalized PMI coordinates are plotted on a triangular graph where the vertices represent prototypical ligand shapes: rod, disc, and sphere (Figure 4A). Upon visual inspection, the libraries populate distinct regions within the triangular plot (Figure 4A, SI Figure 6-1). These observations are supported by the library

averages (SI Table 6-2) and cumulative fraction graphs (Figure 4B, SI Figure 7-1). The Kolmogorov-Smirnov (KS) statistical test and cell-based partitioning were used to quantitatively compare the library distributions (SI-6). The KS test confirmed that the shape distributions of R-BIND (SM) are statistically different from both the FDA and NALDB (SM) libraries, containing more rod-like and less disc- and sphere-like character (SI Table 6-3). Upon partitioning the libraries into four sub-triangles, 72% of R-BIND (SM) populated the “rod” triangle compared to 56% and 52% for the FDA and NALDB (SM) libraries, respectively (SI Table 6-4). Further partitioning into 16 sub-triangles showed that molecules in R-BIND (SM) largely populate the sub-triangles farthest from the sphere vertex (SI Table 6-5). While many libraries are enriched in rod-like molecules,^[23] the differences of R-BIND (SM) compared to the FDA and NALDB (SM) libraries suggest that flat, rod-like small molecules may represent a particularly privileged shape for the recognition of RNA in biological systems.

In summary, we identified properties of biologically active RNA ligands that distinguish them from both RNA binders without reported bioactivity and from FDA-approved drugs that largely target proteins. Specifically, we identified two clear classes of ligands, the monovalent R-BIND (SM) and the multivalent R-BIND (MV), which differ in size and expected binding mode but have both demonstrated RNA target engagement in cell culture and/or animal models. Interestingly, R-BIND (SM) resembled the FDA library in most medicinal chemistry and molecular recognition properties but differed significantly in almost all structural parameters. Notable differences for R-BIND (SM) included a narrow range of positive charge relative to molecular weight, a concurrent increase in nitrogen count and decrease in oxygen count, and differences in ring type and number. The structural properties were also found to be statistically different between the RNA-binding libraries with and without reported bioactivity. Further, the first spatial analysis of R-BIND (SM) small molecules revealed a statistically significant increase in rod-like character of the library and, in particular, a lack of sphere-like shape when compared to other libraries. Together, these differences support the existence of a “privileged chemical space” for RNA targeting and suggest several important future directions, including the explicit testing of these guiding principles against several RNA targets and the expansion of the properties analyzed. Methods including machine-based learning^[24] and multi-parameter optimization^[25] will be explored to design and create libraries^[26] for efficient RNA-focused screening. We expect that our analysis will serve as a stepping stone to facilitate the discovery of chemical probes for RNAs and ultimately facilitate the discovery of small molecule drugs with novel modes of action against humankind’s most challenging diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the members of the Hargrove lab as well as Prof. Hashim Al-Hashimi, Prof. Jiyong Hong, and Prof. David Beratan for stimulating discussion and input. A.E.H. wishes to acknowledge financial support for this work from Duke University, the US National Institute of Health (P50GM103297), and the Prostate Cancer Foundation Young Investigator Award. We also wish to thank: Prof. Agostino Migliore and Dr. Tomasz Janowski for assistance with Molecular Operating Environment Software and analysis of principal moments of inertia calculations; Prof.

Terry Hyslop and Mr. Ren Zhang for assistance with selection, interpretation, and representation of statistical analyses; and Mr. Bilva Sanaba for assistance with compiling and sorting the FDA-approved chemical entities list.

References

1. a) Schmitz SU, Grote P, Herrmann BG. *Cell Mol Life Sci.* 2016; 73:2491. [PubMed: 27007508] b) Scotti MM, Swanson MS. *Nat Rev Genet.* 2016; 17:19. [PubMed: 26593421] c) Chakradhar S. *Nat Med.* 2017; 23:532. [PubMed: 28475574] d) Matsui M, Corey DR. *Nat Rev Drug Discov.* 2017; 16:167. [PubMed: 27444227]
2. Papenfort K, Vogel J. *Cell Host Microbe.* 2010; 8:116. [PubMed: 20638647]
3. a) Tycowski KT, Guo YE, Lee N, Moss WN, Vallery TK, Xie M, Steitz JA. *Genes Dev.* 2015; 29:567. [PubMed: 25792595] b) Fortes P, Morris KV. *Virus Res.* 2016; 212:1. [PubMed: 26454188]
4. a) Ling H, Fabbri M, Calin GA. *Nat Rev Drug Discov.* 2013; 12:847. [PubMed: 24172333] b) Schmitt AM, Chang HY. *Cancer Cell.* 2016; 29:452. [PubMed: 27070700]
5. a) Goff LA, Rinn JL. *Genome Res.* 2015; 25:1456. [PubMed: 26430155] b) Kubota M, Tran C, Spitale RC. *Nat Chem Biol.* 2015; 11:933. [PubMed: 26575240] c) Schlick T, Pyle AM. *Biophys J.* 2017; doi: 10.1016/j.bpj.2016.12.037
6. Garbaccio RM, Parmee ER. *Cell Chem Biol.* 2016; 23:10. [PubMed: 26933732]
7. a) Aboul-ela F. *Future Med Chem.* 2010; 2:93. [PubMed: 21426048] b) Shortridge MD, Varani G. *Curr Opin Struct Biol.* 2015; 30:79. [PubMed: 25687935] c) Disney MD, Angelbello AJ. *Acc Chem Res.* 2016; 49:2698. [PubMed: 27993012]
8. Bodoor K, Boyapati V, Gopu V, Boisdore M, Allam K, Miller J, Treleaven WD, Weldeghiorghis T, Aboul-ela F. *J Med Chem.* 2009; 52:3753. [PubMed: 19445516]
9. Rzuczek SG, Southern MR, Disney MD. *ACS Chem Biol.* 2015; 10:2706. [PubMed: 26414664]
10. a) Bernat V, Disney MD. *Neuron.* 2015; 87:28. [PubMed: 26139368] b) Connelly CM, Moon MH, Schneekloth JS. *Cell Chem Biol.* 2016; 23:1077. [PubMed: 27593111] c) Hermann T. *Wiley Interdiscip Rev : RNA.* 2016; 7:726. [PubMed: 27307213]
11. a) Wilson WD, Li K. *Curr Med Chem.* 2000; 7:73. [PubMed: 10637358] b) Thomas JR, Hergenrother PJ. *Chem Rev.* 2008; 126:224. c) Guan LR, Disney MD. *ACS Chem Biol.* 2012; 7:73. [PubMed: 22185671] d) Le Grice SFJ. *Curr Top Microbiol Immunol.* 2015; 389:147. [PubMed: 25735922]
12. a) Nestor JJ. *Curr Med Chem.* 2009; 16:4399. [PubMed: 19835565] b) Wan WB, Seth PP. *J Med Chem.* 2016; 59:9645. [PubMed: 27434100]
13. Hermann T. *Biopolymers.* 2003; 70:4. [PubMed: 12925990]
14. Palacino J, Swalley SE, Song C, Cheung AK, Shu L, Zhang XL, Van Hoosear M, Shin Y, Chin DN, Keller CG, et al. *Nat Chem Biol.* 2015; 11:511. [PubMed: 26030728]
15. Rzuczek SG, Gao Y, Tang ZZ, Thornton CA, Kodadek T, Disney MD. *ACS Chem Biol.* 2013; 8:2312. [PubMed: 24032410]
16. Wenderski TA, Stratton CF, Bauer RA, Kopp F, Tan DS. *Methods Mol Biol.* 2015; 1263:225. [PubMed: 25618349]
17. Kumar Mishra S, Kumar A. *Database.* 2016; 2016
18. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. *Nucleic Acids Res.* 2006; 34:D668. [PubMed: 16381955]
19. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, et al. *Nat Rev Drug Discov.* 2017; 16:19. [PubMed: 27910877]
20. a) Dandapani S, Marcaurelle LA. *Nat Chem Biol.* 2010; 6:861. [PubMed: 21079589] b) Rafferty RJ, Hicklin RW, Maloof KA, Hergenrother PJ. *Angew Chem Int Ed Engl.* 2014; 53:220. [PubMed: 24273016] *Angew Chem.* 2014; 126:224.
21. a) Stefl R, Skrisovska L, Allain FH. *EMBO Rep.* 2005; 6:33. [PubMed: 15643449] b) Iwakiri J, Tateishi H, Chakraborty A, Patil P, Kenmochi N. *Nucleic Acids Res.* 2012; 40:3299. [PubMed: 22199255] c) Masliah G, Barraud P, Allain FH. *Cell Mol Life Sci.* 2013; 70:1875. [PubMed: 22918483]
22. Sauer WHB, Schwarz MK. *J Chem Inf Comput Sci.* 2003; 43:987. [PubMed: 12767158]

23. Wirth M, Sauer WHB. Mol Inf. 2011; 30:677.
24. a) Mitchell JB. Wiley Interdiscip Rev : Comput Mol Sci. 2014; 4:468. [PubMed: 25285160] b) Lavecchia A. Drug Discov Today. 2015; 20:318. [PubMed: 25448759]
25. Segall MD. Curr Pharm Des. 2012; 18:1292. [PubMed: 22316157]
26. a) Walters WP, Namchuk M. Nat Rev Drug Discov. 2003; 2:259. [PubMed: 12669025] b) Huggins DJ, Venkitaraman AR, Spring DR. ACS Chem Biol. 2011; 6:208. [PubMed: 21261294] c) Dandapani S, Rosse G, Southall N, Salvino JM, Thomas CJ. Curr Protoc Chem Biol. 2012; 4:177. [PubMed: 26705509]

Figure 1.

Examples of library members classified as a) monovalent small molecules (SM, $n = 67$)^[14] and b) multivalent ligands (MV, $n = 37$).^[15]

Figure 2.

Principal component analysis plots based on the cheminformatic parameters calculated for the R-BIND, NALDB, and FDA libraries. a) PCA plot of PC 1 versus PC 2. b) PCA plot of PC 1 versus PC 3. c) PCA plot of PC 2 versus PC 3. d) PCA plot of PC 1-3. The principal component and subsequent percent contribution is indicated on each axis. Outliers were removed here for clarity. Full plots and loading plots can be seen in SI Figure 4-1 and 4-2.

Figure 3.

Box and whisker plots for eight structural parameters identified as statistically different between R-BIND (SM), FDA and NALDB (SM) libraries. The whiskers represent the 10-90th percentile of data, the boxes contain the middle 50% of the data, and the black lines and plus signs denote the medians and means, respectively. Statistically significant differences determined by the Mann Whitney U test are indicated as * $P < 0.05$ and ** $P < 0.001$. N = number of nitrogen atoms; O = number of oxygen atoms; Rings = number of rings; ArRings = number of aromatic rings; HetRings = number of heteroatom-containing rings; SysRings = number of ring systems; Fsp³ = fraction of sp³-hybridized carbons; nStereo = number of stereocenters.

Figure 4.

a) Normalized principal moments of inertia ratios for R-BIND (SM), NALDB (SM) and FDA libraries (140-590 amu). Each point represents the Boltzmann average of a molecule using conformations within 3 kcal/mol of the lowest energy conformer. Larger circles outlined in black represent the library averages. b) Cumulative distance distributions from the rod vertex to members of each library.