IN SILICO RNA APTAMER DESIGN

ALEXANDER WANG*

Abstract. A variety of molecules have been implicated in different physiological and pathophysiological roles. Advances in structural biology have facilitated characterization of the binding domains of many of these molecules, and this has led to discovery and/or development of ligands that can bind a subset of them. However, many of these molecules still lack readily accessible ligands that can reliably be used to selectively bind them and thereby interact with the biological pathways that they are associated with. RNA aptamers are an intriguing modality for such ligands. They offer many advantages, including mobility within biological systems and relatively easy chemical synthesis, but there currently is not a straightforward and unbiased method for generating specific aptamers. Recent advances in parallel computing and scientific machine learning are particularly applicable to the biophysical problem of *in silico* RNA aptamer design. In this paper, we apply parallelization to the problem of searching RNA sequence space for sequence candidates and then model aptameric binding affinity using a neural network whose structure is informed by the sequence-structure relation underlying the biophysics of RNA. We aim to ultimately incorporate both components into a comprehensive computational platform for *de novo* design of RNA aptamers for novel targets.

Key words. parallel computing, scientific machine learning, synthetic biology, biomolecular engineering, RNA design

1. Introduction.

In the field of synthetic biology, there is a significant emphasis on engineering molecules, particularly RNAs and proteins, but the problem of searching the entire sequence space of RNAs or proteins to find a molecule optimized along a certain modality is both experimentally and computationally intractable [55]. Furthermore, while machine learning and other data-driven techniques have been applied to many biological problems, they often ignore the biophysical constraints that govern those systems. In this project, we design and implement a computational pipeline for designing RNA aptamers that is informed by the biophysics underlying aptamer function and stability.

1.1. Diverse Molecular Targets.

Since the completion of the human genome project, innovations in multi-omics technologies and assays have led to the discovery and characterization of a variety of molecular targets involved in various biological mechanisms, such as natural product off-target interactions, and diseases, including cancer and, most recently, COVID-19 [51, 61, 32]. For some of these targets, such as transmembrane G-protein-coupled receptors and kinase families, traditional in vitro screening-based drug discovery has yielded useful binders [4], but effective ligands for others, such as Bcl-2, p53, and RAS, have been more elusive, leaving these targets yet undruggable [60].

To date, discovery and design of such ligands has primarily focused on small molecules and proteins. Small molecules offer advantages in terms of more straightforward development and synthesis, as well as low molecular weight and, therefore, easier penetrance of cells [31], while proteins allows for greater binding specificity and more complex structures and functions [29]. However, more recent characterization of the biophysical properties of ribonucleic acid, or RNA, suggests it as an effective alternative ligand modality.

1.2. RNA Tertiary Structure.

RNA was originally characterized as a single-stranded nucleic acid intermediate for information transfer between DNA and proteins

*Ph.D. Student, Harvard-MIT Program in Health Sciences and Technology, Massachusetts Institute of Technology, Harvard University, Cambridge, MA (azhwang@mit.edu).

but has since been implicated in a variety of other functions, including regulation of gene expression and chromosome maintenance [5, 6], which has spurred interest in its biophysical properties beyond sequence. While most noncoding RNAs (ncRNAs) form RNA-protein complexes, some, such as riboswitches, function via their own structure [8]. Investigation of RNA structure was originally focused on secondary (2-dimensional (2D)) structure, which outlines Watson-Crick base-pairing, but more recently, researchers have begun to place more emphasis on understanding and modeling the tertiary (3-dimensional (3D)) structure of RNA [36].

As understanding of RNA structure and function has grown, researchers have also begun to engineer synthetic RNAs with various functions, including sensing biomolecules and controlling biological systems [58, 16]. In particular, improved understanding of RNA 3D structure is unlocking the potential of engineering a specific class of RNAs: aptamers.

1.3. RNA Aptamers. Nucleic acid aptamers are short, single-stranded oligonucleotides that bind molecular targets with high sensitivity and selectivity, functioning as chemical analogs of antibodies [38]. Because aptamers exhibit 3D structure, they can bind many molecular targets with high affinity, enabling interactions with many biological systems.

1.3.1. Attributes. Compared to antibodies, aptamers are smaller and easier to synthesize [35]. Their smaller size equips them with various advantages, including greater tissue penetration, higher mobility within biological systems, and the ability to interact with a wider range of targets, all of which are particularly relevant in the context of drug development. Their relative ease of chemical synthesis, which does not rely on an immune response, also makes them more practical and cost-efficient to produce and more applicable to widespread use. In fact, since synthesis of aptamers does not require a host, they can be generated for a wider array of targets, including some that may otherwise be toxic to the host [49].

Within nucleic acid aptamers, RNA aptamers also differ from DNA aptamers, with each having certain advantages. DNA aptamers are more chemically stable than RNA aptamers, but the latter exhibit more nuanced three-dimensional structures and thus are more conducive to binding complex targets [63]. Notably, the chemical stability of RNA aptamers can be enhanced by rapid RNA circularization, which is achieved via the Twister-optimized RNA for durable overexpression (Tornado) expression system, to improve their functionality and efficiency in intracellular applications [33].

1.3.2. Structural Motifs. Recurring structural motifs have also been identified within characterized RNA aptamers, including various types of loops (shown in Figure 1), whose secondary and tertiary structures have been implicated in aptamer function, with the majority of binding sites being reported to occur in such loops [30, 22].

1.4. Characterized Aptamers. RNA aptamers have been selected against and generated to bind to a diverse set of molecular targets, including small molecules [9], proteins [53, 28], cells [44], viruses [37], and bacteria [11].

1.4.1. Basic Research. Aptamers that bind various fluorophores and activate their fluorescence have been characterized. The first non-cytotoxic example of these aptamers to be characterized was an analog of green fluorescent protein, named Spinach, which binds the fluorophore 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) [41]. The Spinach aptamer has since been optimized for living

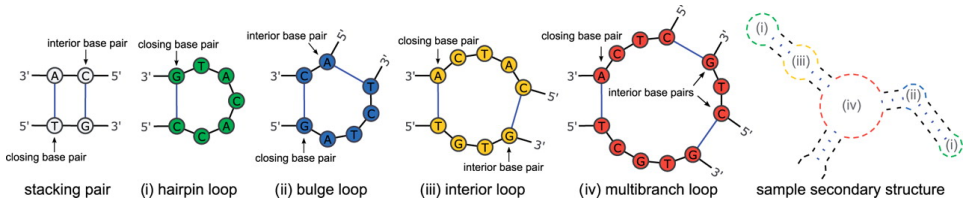

FIG. 1. *Structural motifs in existing aptamers.*

cells to yield the Spinach2 aptamer [46]. Spinach has also been miniaturized to produce the smaller Baby Spinach [57] and Broccoli [20] aptamers, and DFHBI has been optimized for microscopy filter compatibility to form (5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2-methyl-3-(2,2,2-trifluoroethyl)-4H-imidazol-4-one (DFHBI-1T). Researchers have also determined fluorogenic aptamers that amplify other colors, including the RNA Mango aptamers, which bind derivatives of the thiazole orange (TO) fluorophore [14, 3], the complex of which is shown in Figure 2 [50].

1.4.2. Therapeutic and Clinical Examples. RNA aptamers have also been engineered to bind a variety of clinically relevant targets, including viruses, bacteria, and proteins [62]. Pegaptanib, an aptamer that binds the isoform of vascular endothelial growth factor (VEGF) implicated in age-related macular degeneration (AMD), has been approved by the U.S. FDA for treatment of AMD [54]. Another notable clinically relevant target which researchers have selected aptamers against is human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT) [53, 12, 39, 26], the crystal structure of which is shown in Figure 2 [24]. In fact, these aptamers have been shown to bind HIV-1 RT with high affinity, greatly inhibiting its activity, and also with high specificity, as they bind the closely related HIV-2 enzyme with much lower affinity.

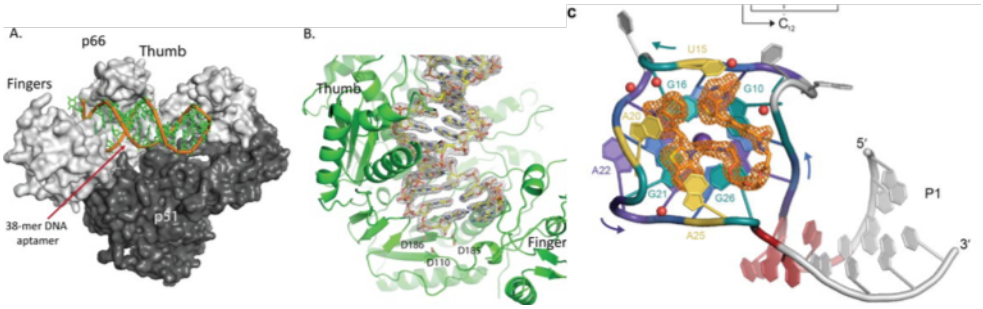

FIG. 2. *Crystal structures of HIV-1 RT aptamer complex (a, b) and Mango-TO1 complex (c).*

1.5. Existing Design Methods. Systematic evolution of ligands by exponential enrichment (SELEX) is an iterative methodology that currently serves as the gold standard for generating aptamers. SELEX involves iteration over binding, partitioning, recovery, and re-amplification steps to select for specific aptamers that are best suited for a preset selection environment [52, 18]. Purified-protein-based SELEX allows for generation of high-affinity aptamers for many protein targets at relatively

low experimental cost but is limited by the target's faithfulness to its native form and function following purification. Whole-cell-based SELEX addresses some of these limitations by allowing selection against certain targets, such as cell surface receptors, in their native environments [42]. Live-animal-based SELEX further addresses these limitations by enabling aptamer selection *in vivo*.

More recently, advances in high-throughput sequencing and bioinformatics have offered deeper insight into individual rounds of the selection process, as opposed to just the final selection landscape, and have also generated large aptamer sequence datasets. However, SELEX still generally returns a single aptamer candidate, which may best satisfy the binding conditions being selected for, but which may be biased and not necessarily constitute the best candidate for therapeutic application [48]. SELEX is also a time-consuming and resource-intensive process. Thus, computational approaches have the potential to augment both the effectiveness and efficiency of targeted aptamer discovery.

1.6. Computational Approach to Engineering RNA. RNA design has long been posited as a computational problem, and for three decades, researchers have developed algorithms for reliably identifying RNA sequences that exhibit specific, desired structures. These algorithms have employed various computational methodologies, including adaptive random walks [21], stochastic local search [2], genetic algorithms [47], Monte Carlo tree search [59], and most recently, reinforcement learning [17]. However, until recently, these algorithms have almost exclusively optimized for secondary structure. Most existing RNA design benchmarks, such as the EteRNA100, also evaluate algorithms based on their ability to predict secondary structure [1]. Researchers have developed one community-sourced benchmarking dataset for RNA tertiary structure, RNA-Puzzles [10, 34], based on an analogous community experiment for protein structure prediction, Critical Assessment of Protein Structure (CASP) [27].

1.6.1. Computational Aptamer Design. Many computational tools have also been constructed for designing aptamers. Initially, researchers drew inspiration from the SELEX procedure and attempted to devise a computational analog of that procedure, resulting in such tools as APTANI [7]. More recently, with the increased spotlight on generative deep learning models, researchers have begun applying this computational paradigm to aptamer design, creating models such as RaptGen [23] and a Restricted Boltzmann Machines-based model [13]. While these models tend to take sequence or secondary structure information as input, they perform modestly well in aptamer discovery.

1.6.2. RNA Tertiary Structure-Based Aptamer Design. The RNA design problem for optimizing for tertiary structure draws many similarities to the secondary structure version. Notably, it likely entails algorithms that also contain a computational model, a fitness function, and a search algorithm [56]. This problem is also similar to the analogous protein design problem, which is focused on the three-dimensional structure of proteins. Just as accurate protein structure prediction by AlphaFold has been a significant boon to *in silico* protein design [25], accurate prediction of RNA tertiary structure will also greatly enhance design of RNAs, particularly in the case of aptamers, whose ability to bind a target is directly a function of its tertiary structure. However, unlike in the case of proteins, for which an abundance of structural data exists in the Protein Data Bank, there is a relative dearth of well-characterized crystal structures of RNAs, but the search space for RNAs is of similarly large magnitude

as that of proteins. This lack of high-resolution training data has thus far been the main obstacle in the way of reliable 3D RNA structure prediction.

1.6.3. RhoFold. Our lab and collaborators have employed various data sampling and augmentation techniques to train RhoFold (formerly named E2EFold-3D), a deep learning model which makes accurate and rapid de novo predictions of RNA three-dimensional structure from sequence [45]. Given the nature of the existing crystal structures that were used for training data, RhoFold is most accurate for predicting structure for short sequences, between 16 and 256 nucleotides, and is thus particularly relevant for RNA aptamers. The performance of existing computational tools for aptamer design, based solely on sequence and/or secondary structure suggests that a model that also considers tertiary structure input, via RhoFold predictions, would be able to produce even more precisely designed aptamers, particularly for existing targets which have well-characterized binding pockets.

2. Parallelized Search of RNA Sequence Space. In the first part of this project, we applied the parallelization principles described in the first part of the course to an RNA aptamer design algorithm.

2.1. RNA Aptamer Design via Sequence Space Search. To design RNA aptamers, we developed an algorithm which explores RNA sequence space, searching for sequences which exhibit a desired three-dimensional structure by using RhoFold to evaluate the three-dimensional structural similarity of candidate sequences to the desired structure. The algorithm is described in Algorithm 2.1.

Algorithm 2.1 Search of RNA Sequence Space Starting from Given Sequence

while search has not converged and max iterations not reached **do**
  for each position i in the current sequence s **do**
    for each base b in the set $\{A, U, C, G\} \cap \{s[i]\}$ **do**
      Replace (mutate) $s[i]$ with b to construct the sequence s'
      Use RhoFold to predict the tertiary structure of s'
      Compute and store the structural difference between s' and the target
    end for
  end for
  Apply the mutation which yielded the closest structure to the target to s
  If no mutations improve the structural similarity to the target, end search
end while

Essentially, the algorithm takes an initial sequence and performs a succession of local searches. At each iteration, it considers all sequences that are a single-base mutation away and, in aggregate, constructs a path of the single-base mutations which improve the tertiary structural similarity to the target structure the most, as visualized in Figure 3 [43].

Unfortunately, RhoFold [45] and the current structural similarity function are both implemented in Python, and the PyCall package for calling Python functions in Julia does not work with multi-threading in Julia, so to still demonstrate the parallelization in this framework and evaluate its performance, I used two simple Julia functions to simulate tertiary structure prediction and structural similarity computation.

2.2. Parallelization Performance. While the main loop of the algorithm continually updates the current sequence being considered and is not conducive to par-

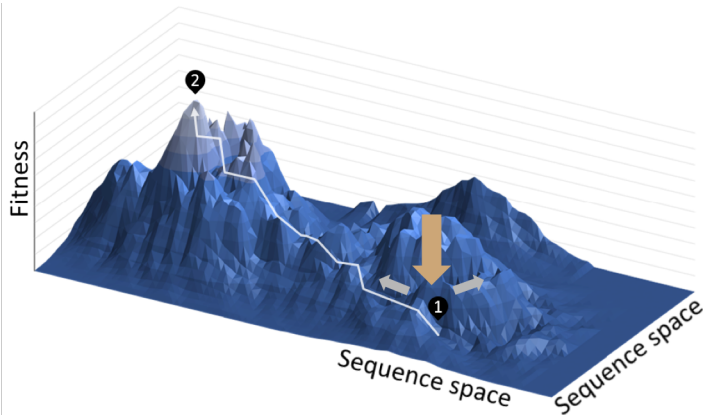

FIG. 3. *Successive local searches from a single starting sequence.*

allelization, the local search within each iteration of all single-base mutations of the current sequence can be parallelized, as consideration of each mutation does not affects that of others. The performances of the serial and parallel versions of this algorithm are described in Table 1.

TABLE 1
Benchmarks for serial and parallel versions of single starting sequence algorithm.

Threads	1	16
Range (min ... max)	3.552 ms ... 8.892 ms	948.655 μs ... 41.415 ms
Time (median)	4.238 ms	2.195 ms
Time (mean $\pm$ σ)	4.260 ms $\pm$ 521.851 μs	2.752 ms $\pm$ 3.163 ms

2.3. Further Parallelization. We also modified the overall sequence space search algorithm to further apply parallelization by adding an additional outer loop to consider multiple starting sequences simultaneously, as described in Algorithm 2.2 and visualized in Figure 4 [43]. Since the single-base mutation path sought for each starting sequence does not affect or depend on that of any other starting sequence, this outer loop is also conducive to parallelization.

Algorithm 2.2 Search of RNA Sequence Space Starting from Multiple Sequences

```
for each starting sequence (parallelized) do
  while search has not converged and max iterations not reached do
    for each position i in the current sequence s (parallelized) do
      for each base b in the set {A, U, C, G} ∩ {s[i]} do
        Replace (mutate) s[i] with b to construct the sequence s′
        Use RhoFold to predict the tertiary structure of s′
        Compute and store the structural difference between s′ and the target
      end for
    end for
    Apply the mutation which yielded the closest structure to the target to s
    If no mutations improve the structural similarity to the target, end search
  end while
end for
```

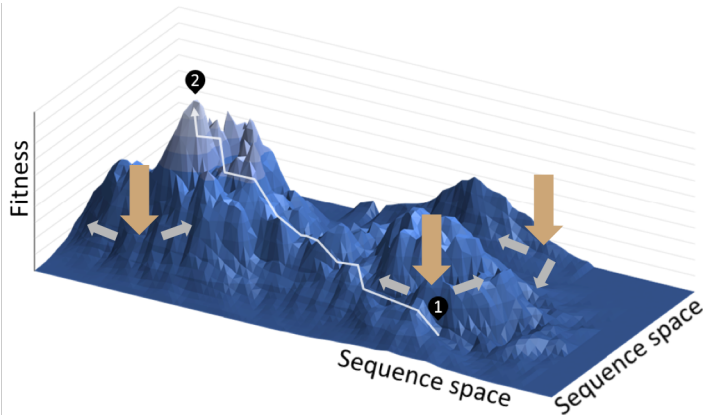

FIG. 4. *Successive local searches from a single starting sequence.*

The performances of the serial and parallel versions of this improved algorithm are described in Table 2.

TABLE 2
Benchmarks for serial and parallel versions of multiple starting sequence algorithm.

Threads	1	16
Range (min ... max)	10.625 ms ... 17.991 ms	1.987 ms ... 172.053 ms
Time (median)	12.561 ms	3.244 ms
Time (mean $\pm$ σ)	12.491 ms $\pm$ 1.072 ms	5.581 ms $\pm$ 10.064 ms

2.4. Discussion. For the algorithm which conducted the search from only a single starting sequence, parallelization with 16 threads resulted in about a 1.5× to 2× speedup. For the algorithm which conducted the search from multiple starting sequences, parallelization with 16 threads resulted in about a 2× to 4× speedup. Notably, the range of the multi-threaded algorithm runtime is much wider than that of the single-threaded algorithm, suggesting that there are certain cases in which multi-threading is actually detrimental to the performance of the algorithm. In the same vein, using 16 threads does not produce a 16× speedup, but the speedup is instead more muted, although still very apparent.

Although the structural prediction and similarity functions used in this analysis are much simpler than the true functions, this speedup via parallelization should translate to when using the true functions as well.

3. Scientific Machine Learning to Model Aptamer Function. In the second part of this project, we applied the scientific machine learning principles described in the second part of this course to model the aptameric binding affinity data generated by ordering and experimentally testing the sequence candidates identified in the first part of the project.

3.1. Biophysics-Informed RNA Sequence Input Encoding. To build this binding affinity model, we employed a technique inspired by a previously described approach to protein design [40]. We used the sequence-structure relation underlying the biophysics of RNAs to inform the structure of our input and the hidden layers of our neural network. Specifically, we encoded our RNA sequence input as a one hot embedding at each residue, as each residue must be one of four bases (A, U, C, G),

resulting in a $4 \times n$ matrix input, where n is the length of the aptamer/sequence we are modeling. We then unraveled this matrix into a $4n$-vector to use as the input to our neural network.

While this current implementation did not necessarily follow the formal physics-informed neural network paradigm described in class leveraging partial differential equations (PDEs) governing the system being modeled, it does follow a broader scientific machine learning paradigm wherein it incorporates biophysical characteristics of this system into the training process and construction of the model. Our current data also only includes overall binding affinity. Upon expanding our data collection and methodology to model the molecular docking interactions between aptamers and their targets driving this binding affinity, we will be able to employ true physics-informed neural networks governed by the PDE systems underlying such molecular docking interactions.

3.2. Neural Network Architecture. We designed our neural network to have the architecture described in Figure 5 to attempt to learn specific biophysically-rooted relationships. Specifically, we used two sets of weights in our model, with the first aimed at learning relationships between specific individual residues in the sequence and the second aimed at learning more complex relationships between larger domains of multiple residues identified via the first. Figure 5 shows an example architecture of the network that would be used to model a sequence of length 3.

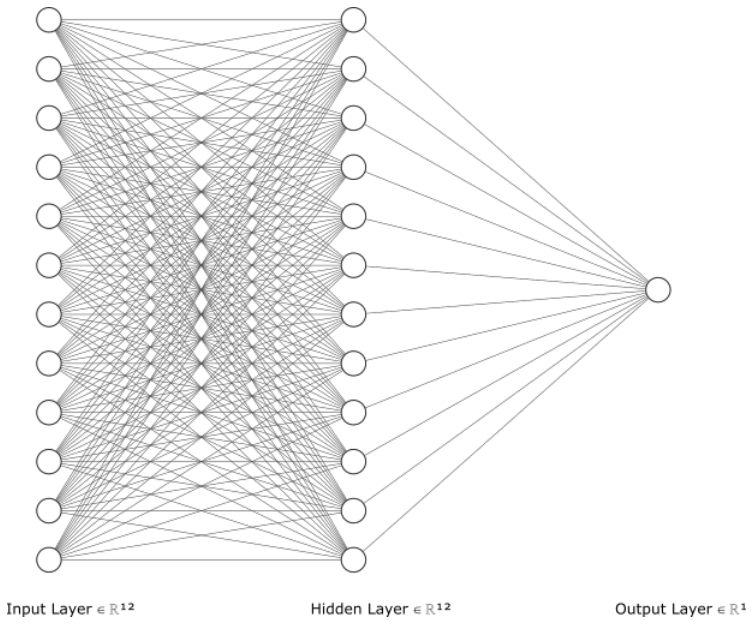

FIG. 5. *Neural network architecture.*

3.3. RNA Mango Results. We first trained our neural network on binding affinity data obtained by experimentally testing our first set of RNA Mango aptamer candidates that we generated in the first part of the project. We can see from Figure 6 that the model converges pretty quickly.

Since our dataset was small, we evaluated the performance of our neural network by examining its edge weights. Specifically, we found the edge weights to modestly

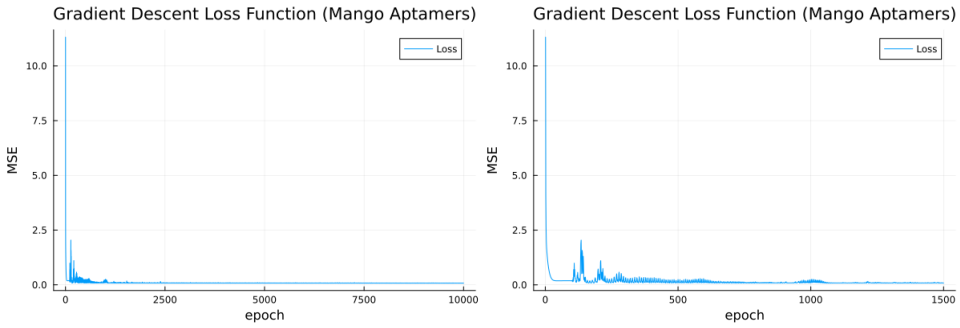

FIG. 6. *Loss function for neural network trained on RNA Mango aptamer data.*

emphasize certain motifs previously identified in literature as playing important roles in the original Mango aptamer's function [3].

3.4. HIV RT Results. We then trained a neural network with the same architecture (Figure 5) on binding affinity data obtained by experimentally testing our first set of HIV RT aptamer candidates that we also generated in the first part of the project. We can see from Figure 7 that the model again converges pretty quickly.

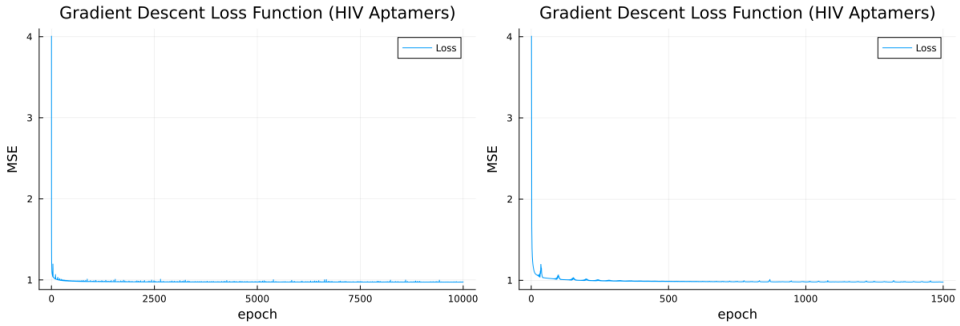

FIG. 7. *Loss function for neural network trained on HIV RT aptamer data.*

Due to the small size of our dataset, we again evaluated our model by considering its edge weights and again found modest emphasis on motifs reported in literature as important in HIV RT aptamer function [53]. Notably, these HIV RT aptamer motifs were more heavily weighted than those emphasized in the Mango aptamer model.

3.5. Discussion. Since both the Mango aptamer and HIV RT aptamer experimental datasets were relatively small, the models trained on them likely vastly overfit the data, decreasing their effectiveness at designing novel, dissimilar aptamers that recapitulate the respective binding affinities. However, we did find the biophysics-informed design of the model to allow us to identify certain motifs known to be important to aptamer function, which suggests that, with more data, this methodology could be used to identify additional novel motifs involved in aptamer binding affinity.

4. Conclusions. RNA aptamers present an interesting modality with which to design ligands for binding many currently undruggable biological targets. However, there remains a need for a reliable methodology for designing these aptamers for

specific targets *de novo*. In this study, we applied parallelization to optimize the performance of an RNA design algorithm which works by searching RNA sequence space and identifying candidate sequences by evaluating their predicted structural similarity. We then applied a scientific machine learning framework to model the binding affinities of candidate sequences by training a neural network whose structure was informed by RNA sequence-structure biophysics on experimental data. Both of these computational methodologies will be integrated into an overarching general RNA aptamer design algorithm, which we ultimately aim to use the engine to power an RNA aptamer design pipeline, described in Figure 8, that will be able to take as input any biological or chemical target with a well-characterized binding domain and design an RNA aptamer which binds it with high specificity and affinity.

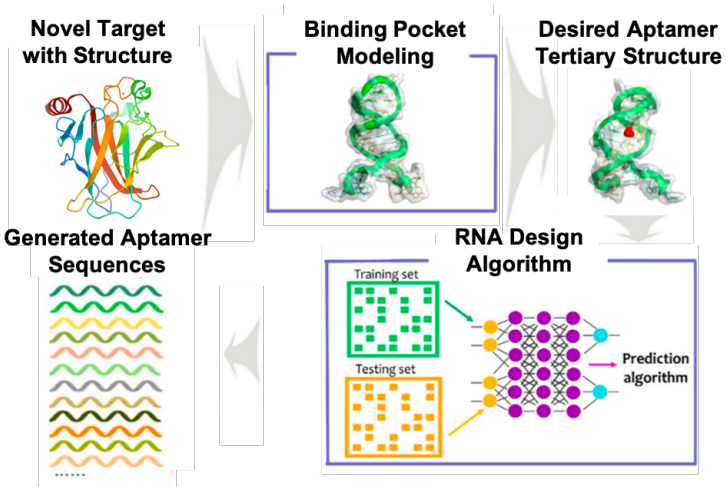

FIG. 8. *Proposed RNA aptamer design pipeline. Novel target crystal structure (human mutant p53) and pipeline figure adapted from [19] and [15], respectively.*

5. Code Availability. A Jupyter Notebook containing the code described in this paper can be found at https://github.com/azhwang5/parallel-computing-sciml-rna-aptamer-design.

REFERENCES

[1] J. ANDERSON-LEE, E. FISKER, V. KOSARAJU, M. WU, J. KONG, J. LEE, M. LEE, M. ZADA, A. TREUILLE, AND R. DAS, *Principles for Predicting RNA Secondary Structure Design Difficulty*, Journal of Molecular Biology, 428 (2016), pp. 748–757, https://doi.org/10.1016/j.jmb.2015.11.013, https://linkinghub.elsevier.com/retrieve/pii/S0022283615006567 (accessed 2023-04-25).

[2] M. ANDRONESCU, A. P. FEJES, F. HUTTER, H. H. HOOS, AND A. CONDON, *A New Algorithm for RNA Secondary Structure Design*, Journal of Molecular Biology, 336 (2004), pp. 607–624, https://doi.org/10.1016/j.jmb.2003.12.041, https://linkinghub.elsevier.com/retrieve/pii/S0022283603015596 (accessed 2023-04-25).

[3] A. AUTOUR, S. C. Y. JENG, A. D. CAWTE, A. ABDOLAHZADEH, A. GALLI, S. S. S. PANCHAPAKESAN, D. RUEDA, M. RYCKELYNCK, AND P. J. UNRAU, *Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells*, Nature Communications, 9 (2018), p. 656, https://doi.org/10.1038/s41467-018-02993-8, https://www.nature.com/articles/s41467-018-02993-8 (accessed 2023-04-25).

[4] M. L. Billingsley, *Druggable Targets and Targeted Drugs: Enhancing the Development of New Therapeutics*, Pharmacology, 82 (2008), pp. 239–244, https://doi.org/10.1159/000157624, https://www.karger.com/Article/FullText/157624 (accessed 2023-04-25).

[5] S. Brenner, F. Jacob, and M. Meselson, *An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis*, Nature, 190 (1961), pp. 576–581, https://doi.org/10.1038/190576a0, https://www.nature.com/articles/190576a0 (accessed 2023-04-25).

[6] M. G. Caprara and T. W. Nilsen, *RNA: versatility in form and function*, Nature Structural Biology, 7 (2000), pp. 831–833, https://doi.org/10.1038/82816, http://www.nature.com/doifinder/10.1038/82816 (accessed 2023-04-25).

[7] J. Caroli, C. Taccioli, A. De La Fuente, P. Serafini, and S. Bicciato, *APTANI: a computational tool to select aptamers through sequence-structure motif analysis of HT-SELEX data*, Bioinformatics, 32 (2016), pp. 161–164, https://doi.org/10.1093/bioinformatics/btv545, https://academic.oup.com/bioinformatics/article/32/2/161/1743600 (accessed 2023-04-25).

[8] T. Cech and J. Steitz, *The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones*, Cell, 157 (2014), pp. 77–94, https://doi.org/10.1016/j.cell.2014.03.008, https://linkinghub.elsevier.com/retrieve/pii/S0092867414003389 (accessed 2023-04-25).

[9] J. Ciesiolka and M. Yarus, *Small RNA-divalent domains*, RNA (New York, N.Y.), 2 (1996), pp. 785–793.

[10] J. A. Cruz, M.-F. Blanchet, M. Boniecki, J. M. Bujnicki, S.-J. Chen, S. Cao, R. Das, F. Ding, N. V. Dokholyan, S. C. Flores, L. Huang, C. A. Lavender, V. Lisi, F. Major, K. Mikolajczak, D. J. Patel, A. Philips, T. Puton, J. Santalucia, F. Sijenyi, T. Hermann, K. Rother, M. Rother, A. Serganov, M. Skorupski, T. Soltysinski, P. Sripakdeevong, I. Tuszynska, K. M. Weeks, C. Waldsich, M. Wildauer, N. B. Leontis, and E. Westhof, *RNA-Puzzles : A CASP-like evaluation of RNA three-dimensional structure prediction*, RNA, 18 (2012), pp. 610–625, https://doi.org/10.1261/rna.031054.111, http://rnajournal.cshlp.org/lookup/doi/10.1261/rna.031054.111 (accessed 2023-04-25).

[11] A. S. Davydova, M. A. Vorobyeva, M. R. Kabilov, N. V. Tikunova, D. V. Pyshnyi, and A. G. Venyaminova, *In vitro selection of cell-internalizing 2'-modified RNA aptamers against Pseudomonas aeruginosa*, Russian Journal of Bioorganic Chemistry, 43 (2017), pp. 58–63, https://doi.org/10.1134/S1068162016060030, http://link.springer.com/10.1134/S1068162016060030 (accessed 2023-04-25).

[12] A. D. Dearborn, E. Eren, N. R. Watts, I. W. Palmer, J. D. Kaufman, A. C. Steven, and P. T. Wingfield, *Structure of an RNA Aptamer that Can Inhibit HIV-1 by Blocking Rev-Cognate RNA (RRE) Binding and Rev-Rev Association*, Structure, 26 (2018), pp. 1187–1195.e4, https://doi.org/10.1016/j.str.2018.06.001, https://linkinghub.elsevier.com/retrieve/pii/S0969212618302077 (accessed 2023-04-25).

[13] A. Di Gioacchino, J. Procyk, M. Molari, J. S. Schreck, Y. Zhou, Y. Liu, R. Monasson, S. Cocco, and P. Šulc, *Generative and interpretable machine learning for aptamer design and analysis of in vitro sequence selection*, PLOS Computational Biology, 18 (2022), p. e1010561, https://doi.org/10.1371/journal.pcbi.1010561, https://dx.plos.org/10.1371/journal.pcbi.1010561 (accessed 2023-04-25).

[14] E. V. Dolgosheina, S. C. Y. Jeng, S. S. S. Panchapakesan, R. Cojocaru, P. S. K. Chen, P. D. Wilson, N. Hawkins, P. A. Wiggins, and P. J. Unrau, *RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking*, ACS Chemical Biology, 9 (2014), pp. 2412–2420, https://doi.org/10.1021/cb500499x, https://pubs.acs.org/doi/10.1021/cb500499x (accessed 2023-04-25).

[15] A. Douaki, D. Garoli, A. K. M. S. Inam, M. A. C. Angeli, G. Cantarella, W. Rocchia, J. Wang, L. Petti, and P. Lugli, *Smart Approach for the Design of Highly Selective Aptamer-Based Biosensors*, Biosensors, 12 (2022), p. 574, https://doi.org/10.3390/bios12080574, https://www.mdpi.com/2079-6374/12/8/574 (accessed 2023-04-26).

[16] P. B. Dykstra, M. Kaplan, and C. D. Smolke, *Engineering synthetic RNA devices for cell control*, Nature Reviews Genetics, 23 (2022), pp. 215–228, https://doi.org/10.1038/s41576-021-00436-7, https://www.nature.com/articles/s41576-021-00436-7 (accessed 2023-04-25).

[17] P. Eastman, J. Shi, B. Ramsundar, and V. S. Pande, *Solving the RNA design problem with reinforcement learning*, PLOS Computational Biology, 14 (2018), p. e1006176, https://doi.org/10.1371/journal.pcbi.1006176, https://dx.plos.org/10.1371/journal.pcbi.1006176 (accessed 2023-04-25).

[18] A. D. Ellington and J. W. Szostak, *In vitro selection of RNA molecules that bind specific*

ligands, Nature, 346 (1990), pp. 818–822, https://doi.org/10.1038/346818a0, http://www.nature.com/articles/346818a0 (accessed 2023-04-25).

[19] S. Emamzadah, L. Tropia, I. Vincenti, B. Falquet, and T. D. Halazonetis, *Reversal of the DNA-Binding-Induced Loop L1 Conformational Switch in an Engineered Human p53 Protein*, Journal of Molecular Biology, 426 (2014), pp. 936–944, https://doi.org/10.1016/j.jmb.2013.12.020, https://linkinghub.elsevier.com/retrieve/pii/S0022283613008073 (accessed 2023-04-26).

[20] G. S. Filonov, J. D. Moon, N. Svensen, and S. R. Jaffrey, *Broccoli: Rapid Selection of an RNA Mimic of Green Fluorescent Protein by Fluorescence-Based Selection and Directed Evolution*, Journal of the American Chemical Society, 136 (2014), pp. 16299–16308, https://doi.org/10.1021/ja508478x, https://pubs.acs.org/doi/10.1021/ja508478x (accessed 2023-04-25).

[21] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster, *Fast folding and comparison of RNA secondary structures*, Monatshefte für Chemie Chemical Monthly, 125 (1994), pp. 167–188, https://doi.org/10.1007/BF00818163, http://link.springer.com/10.1007/BF00818163 (accessed 2023-04-25).

[22] J. Hoinka, E. Zotenko, A. Friedman, Z. E. Sauna, and T. M. Przytycka, *Identification of sequence–structure RNA binding motifs for SELEX-derived aptamers*, Bioinformatics, 28 (2012), pp. i215–i223, https://doi.org/10.1093/bioinformatics/bts210, https://academic.oup.com/bioinformatics/article/28/12/i215/267900 (accessed 2023-05-11).

[23] N. Iwano, T. Adachi, K. Aoki, Y. Nakamura, and M. Hamada, *Generative aptamer discovery using RaptGen*, Nature Computational Science, 2 (2022), pp. 378–386, https://doi.org/10.1038/s43588-022-00249-6, https://www.nature.com/articles/s43588-022-00249-6 (accessed 2023-04-25).

[24] J. Jaeger, *The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor*, The EMBO Journal, 17 (1998), pp. 4535–4542, https://doi.org/10.1093/emboj/17.15.4535, http://emboj.embopress.org/cgi/doi/10.1093/emboj/17.15.4535 (accessed 2023-04-25).

[25] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, *Highly accurate protein structure prediction with AlphaFold*, Nature, 596 (2021), pp. 583–589, https://doi.org/10.1038/s41586-021-03819-2, https://www.nature.com/articles/s41586-021-03819-2 (accessed 2023-04-25).

[26] O. Kensch, B. A. Connolly, H.-J. Steinhoff, A. McGregor, R. S. Goody, and T. Restle, *HIV-1 Reverse Transcriptase-Pseudoknot RNA Aptamer Interaction Has a Binding Affinity in the Low Picomolar Range Coupled with High Specificity*, Journal of Biological Chemistry, 275 (2000), pp. 18271–18278, https://doi.org/10.1074/jbc.M001309200, https://linkinghub.elsevier.com/retrieve/pii/S0021925819831824 (accessed 2023-04-25).

[27] A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, and J. Moult, *Critical assessment of methods of protein structure prediction (CASP)—Round XIV*, Proteins: Structure, Function, and Bioinformatics, 89 (2021), pp. 1607–1617, https://doi.org/10.1002/prot.26237, https://onlinelibrary.wiley.com/doi/10.1002/prot.26237 (accessed 2023-04-25).

[28] M. F. Kubik, A. W. Stephens, D. Schneider, R. A. Marlar, and D. Tasset, *High-affinity RNA ligands to human α-thrombin*, Nucleic Acids Research, 22 (1994), pp. 2619–2626, https://doi.org/10.1093/nar/22.13.2619, https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/22.13.2619 (accessed 2023-04-25).

[29] B. Leader, Q. J. Baca, and D. E. Golan, *Protein therapeutics: a summary and pharmacological classification*, Nature Reviews Drug Discovery, 7 (2008), pp. 21–39, https://doi.org/10.1038/nrd2399, https://www.nature.com/articles/nrd2399 (accessed 2023-04-25).

[30] J. F. Lee, *Aptamer Database*, Nucleic Acids Research, 32 (2004), pp. 95D–100, https://doi.org/10.1093/nar/gkh094, https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkh094 (accessed 2023-05-11).

[31] Q. Li and C. Kang, *Mechanisms of Action for Small Molecules Revealed by Structural Biology in Drug Discovery*, International Journal of Molecular Sciences, 21 (2020), p. 5262, https://doi.org/10.3390/ijms21155262, https://www.mdpi.com/1422-0067/21/15/5262 (accessed 2023-04-25).

[32] Y. Li, G. Hou, H. Zhou, Y. Wang, H. M. Tun, A. Zhu, J. Zhao, F. Xiao, S. Lin, D. Liu, D. Zhou, L. Mai, L. Zhang, Z. Zhang, L. Kuang, J. Guan, Q. Chen, L. Wen, Y. Zhang, J. Zhuo, F. Li, Z. Zhuang, Z. Chen, L. Luo, D. Liu, C. Chen, M. Gan, N. Zhong, J. Zhao, Y. Ren, and Y. Xu, *Multi-platform omics analysis reveals molecular signature for COVID-19 pathogenesis, prognosis and drug target discovery*, Signal Transduction and Targeted Therapy, 6 (2021), p. 155, https://doi.org/10.1038/s41392-021-00508-4, https://www.nature.com/articles/s41392-021-00508-4 (accessed 2023-04-25).

[33] J. L. Litke and S. R. Jaffrey, *Highly efficient expression of circular RNA aptamers in cells using autocatalytic transcripts*, Nature Biotechnology, 37 (2019), pp. 667–675, https://doi.org/10.1038/s41587-019-0090-6, http://www.nature.com/articles/s41587-019-0090-6 (accessed 2023-04-25).

[34] M. Magnus, M. Antczak, T. Zok, J. Wiedemann, P. Lukasiak, Y. Cao, J. M. Bujnicki, E. Westhof, M. Szachniuk, and Z. Miao, *RNA-Puzzles toolkit: a computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools*, Nucleic Acids Research, 48 (2020), pp. 576–588, https://doi.org/10.1093/nar/gkz1108.

[35] G. Mayer, *The Chemical Biology of Aptamers*, Angewandte Chemie International Edition, 48 (2009), pp. 2672–2689, https://doi.org/10.1002/anie.200804643, https://onlinelibrary.wiley.com/doi/10.1002/anie.200804643 (accessed 2023-04-25).

[36] Z. Miao and E. Westhof, *RNA Structure: Advances and Assessment of 3D Structure Prediction*, Annual Review of Biophysics, 46 (2017), pp. 483–503, https://doi.org/10.1146/annurev-biophys-070816-034125, https://www.annualreviews.org/doi/10.1146/annurev-biophys-070816-034125 (accessed 2023-04-25).

[37] T. S. Misono and P. K. Kumar, *Selection of RNA aptamers against human influenza virus hemagglutinin using surface plasmon resonance*, Analytical Biochemistry, 342 (2005), pp. 312–317, https://doi.org/10.1016/j.ab.2005.04.013, https://linkinghub.elsevier.com/retrieve/pii/S000326970500309X (accessed 2023-04-25).

[38] X. Ni, M. Castanares, A. Mukherjee, and S. Lupold, *Nucleic Acid Aptamers: Clinical Applications and Promising New Horizons*, Current Medicinal Chemistry, 18 (2011), pp. 4206–4214, https://doi.org/10.2174/092986711797189600, http://www.eurekaselect.com/openurl/content.php?genre=article&issn=0929-8673&volume=18&issue=27&spage=4206 (accessed 2023-04-25).

[39] D. G. Nickens, J. T. Patterson, and D. H. Burke, *Inhibition of HIV-1 reverse transcriptase by RNA aptamers in Escherichia coli*, RNA, 9 (2003), pp. 1029–1033, https://doi.org/10.1261/rna.5550103, http://rnajournal.cshlp.org/lookup/doi/10.1261/rna.5550103 (accessed 2023-04-25).

[40] S. I. Omar, C. Keasar, A. J. Ben-Sasson, and E. Haber, *Protein Design Using Physics Informed Neural Networks*, Biomolecules, 13 (2023), p. 457, https://doi.org/10.3390/biom13030457, https://www.mdpi.com/2218-273X/13/3/457 (accessed 2023-05-16).

[41] J. S. Paige, K. Y. Wu, and S. R. Jaffrey, *RNA Mimics of Green Fluorescent Protein*, Science, 333 (2011), pp. 642–646, https://doi.org/10.1126/science.1207339, https://www.science.org/doi/10.1126/science.1207339 (accessed 2023-04-25).

[42] K. Sefah, D. Shangguan, X. Xiong, M. B. O'Donoghue, and W. Tan, *Development of DNA aptamers using Cell-SELEX*, Nature Protocols, 5 (2010), pp. 1169–1185, https://doi.org/10.1038/nprot.2010.66, https://www.nature.com/articles/nprot.2010.66 (accessed 2023-04-25).

[43] T. Shafee, *Evolvability of a viral protease: experimental evolution of catalysis, robustness and specificity*, (2014), https://doi.org/10.17863/CAM.16528, https://www.repository.cam.ac.uk/handle/1810/245207 (accessed 2023-04-26). Publisher: Apollo - University of Cambridge Repository.

[44] D. Shangguan, L. Meng, Z. C. Cao, Z. Xiao, X. Fang, Y. Li, D. Cardona, R. P. Witek, C. Liu, and W. Tan, *Identification of Liver Cancer-Specific Aptamers Using Whole Live Cells*, Analytical Chemistry, 80 (2008), pp. 721–728, https://doi.org/10.1021/ac701962v, https://pubs.acs.org/doi/10.1021/ac701962v (accessed 2023-04-25).

[45] T. Shen, Z. Hu, Z. Peng, J. Chen, P. Xiong, L. Hong, L. Zheng, Y. Wang, I. King, S. Wang, S. Sun, and Y. Li, *E2Efold-3D: End-to-End Deep Learning Method for accurate de novo RNA 3D Structure Prediction*, July 2022, http://arxiv.org/abs/2207.01586 (accessed 2023-04-25). arXiv:2207.01586 [cs, q-bio].

[46] R. L. Strack, M. D. Disney, and S. R. Jaffrey, *A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat–containing RNA*, Nature Methods, 10 (2013), pp. 1219–1224, https://doi.org/10.1038/nmeth.2701, http://www.nature.com/articles/nmeth.2701 (accessed 2023-04-25).

[47] A. Taneda, *MODENA: a multi-objective RNA inverse folding*, Ad-

vances and Applications in Bioinformatics and Chemistry, (2010), p. 1, https://doi.org/10.2147/AABC.S14335, http://www.dovepress.com/modena-a-multi-objective-rna-inverse-folding-peer-reviewed-article-AABC (accessed 2023-04-25).

[48] W. H. Thiel, T. Bair, K. Wyatt Thiel, J. P. Dassie, W. M. Rockey, C. A. Howell, X. Y. Liu, A. J. Dupuy, L. Huang, R. Owczarzy, M. A. Behlke, J. O. McNamara, and P. H. Giangrande, *Nucleotide Bias Observed with a Short SELEX RNA Aptamer Library*, Nucleic Acid Therapeutics, 21 (2011), pp. 253–263, https://doi.org/10.1089/nat.2011.0288, http://www.liebertpub.com/doi/10.1089/nat.2011.0288 (accessed 2023-04-25).

[49] V. Thiviyanathan and D. G. Gorenstein, *Aptamers and the next generation of diagnostic reagents*, PROTEOMICS - Clinical Applications, 6 (2012), pp. 563–573, https://doi.org/10.1002/prca.201200042, https://onlinelibrary.wiley.com/doi/10.1002/prca.201200042 (accessed 2023-04-25).

[50] R. J. Trachman, N. A. Demeshkina, M. W. L. Lau, S. S. S. Panchapakesan, S. C. Y. Jeng, P. J. Unrau, and A. R. Ferré-D'Amaré, *Structural basis for high-affinity fluorophore binding and activation by RNA Mango*, Nature Chemical Biology, 13 (2017), pp. 807–813, https://doi.org/10.1038/nchembio.2392, http://www.nature.com/articles/nchembio.2392 (accessed 2023-05-11).

[51] B. Tran, J. E. Dancey, S. Kamel-Reid, J. D. McPherson, P. L. Bedard, A. M. Brown, T. Zhang, P. Shaw, N. Onetto, L. Stein, T. J. Hudson, B. G. Neel, and L. L. Siu, *Cancer Genomics: Technology, Discovery, and Translation*, Journal of Clinical Oncology, 30 (2012), pp. 647–660, https://doi.org/10.1200/JCO.2011.39.2316, https://ascopubs.org/doi/10.1200/JCO.2011.39.2316 (accessed 2023-04-25).

[52] C. Tuerk and L. Gold, *Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase*, Science, 249 (1990), pp. 505–510, https://doi.org/10.1126/science.2200121, https://www.science.org/doi/10.1126/science.2200121 (accessed 2023-04-25).

[53] C. Tuerk, S. MacDougal, and L. Gold, *RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase.*, Proceedings of the National Academy of Sciences, 89 (1992), pp. 6988–6992, https://doi.org/10.1073/pnas.89.15.6988, https://pnas.org/doi/full/10.1073/pnas.89.15.6988 (accessed 2023-04-25).

[54] S. A. Vinores, *Pegaptanib in the treatment of wet, age-related macular degeneration*, International Journal of Nanomedicine, 1 (2006), pp. 263–268.

[55] C. A. Voigt, S. L. Mayo, F. H. Arnold, and Z.-G. Wang, *Computational method to reduce the search space for directed protein evolution*, Proceedings of the National Academy of Sciences, 98 (2001), pp. 3778–3783, https://doi.org/10.1073/pnas.051614498, https://pnas.org/doi/full/10.1073/pnas.051614498 (accessed 2023-05-13).

[56] M. Ward, E. Courtney, and E. Rivas, *Fitness functions for RNA structure design*, Nucleic Acids Research, 51 (2023), pp. e40–e40, https://doi.org/10.1093/nar/gkad097, https://academic.oup.com/nar/article/51/7/e40/7068363 (accessed 2023-04-25).

[57] K. D. Warner, M. C. Chen, W. Song, R. L. Strack, A. Thorn, S. R. Jaffrey, and A. R. Ferré-D'Amaré, *Structural basis for activity of highly efficient RNA mimics of green fluorescent protein*, Nature Structural & Molecular Biology, 21 (2014), pp. 658–663, https://doi.org/10.1038/nsmb.2865, http://www.nature.com/articles/nsmb.2865 (accessed 2023-04-25).

[58] Y. Xiu, S. Jang, J. A. Jones, N. A. Zill, R. J. Linhardt, Q. Yuan, G. Y. Jung, and M. A. G. Koffas, *Naringenin-responsive riboswitch-based fluorescent biosensor module for Escherichia coli co-cultures: Use of an Aptasensor for Flavonoid Screening in E.coli Co-culture*, Biotechnology and Bioengineering, 114 (2017), pp. 2235–2244, https://doi.org/10.1002/bit.26340, https://onlinelibrary.wiley.com/doi/10.1002/bit.26340 (accessed 2023-04-25).

[59] X. Yang, K. Yoshizoe, A. Taneda, and K. Tsuda, *RNA inverse folding using Monte Carlo tree search*, BMC Bioinformatics, 18 (2017), p. 468, https://doi.org/10.1186/s12859-017-1882-7, https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1882-7 (accessed 2023-04-25).

[60] G. Zhang, J. Zhang, Y. Gao, Y. Li, and Y. Li, *Strategies for targeting undruggable targets*, Expert Opinion on Drug Discovery, 17 (2022), pp. 55–69, https://doi.org/10.1080/17460441.2021.1969359, https://www.tandfonline.com/doi/full/10.1080/17460441.2021.1969359 (accessed 2023-04-25).

[61] H.-W. Zhang, C. Lv, L.-J. Zhang, X. Guo, Y.-W. Shen, D. G. Nagle, Y.-D. Zhou, S.-H. Liu, W.-D. Zhang, and X. Luan, *Application of omics- and multi-omics-based techniques for natural product target discovery*, Biomedicine & Pharmacotherapy, 141 (2021),

p. 111833, https://doi.org/10.1016/j.biopha.2021.111833, https://linkinghub.elsevier.com/retrieve/pii/S0753332221006156 (accessed 2023-04-25).

[62] J. Zhou and J. Rossi, *Aptamers as targeted therapeutics: current potential and challenges*, Nature Reviews Drug Discovery, 16 (2017), pp. 181–202, https://doi.org/10.1038/nrd.2016.199, http://www.nature.com/articles/nrd.2016.199 (accessed 2023-04-25).

[63] J. Zhou and J. J. Rossi, *Cell-type-specific, Aptamer-functionalized Agents for Targeted Disease Therapy*, Molecular Therapy - Nucleic Acids, 3 (2014), p. e169, https://doi.org/10.1038/mtna.2014.21, https://linkinghub.elsevier.com/retrieve/pii/S2162253116303092 (accessed 2023-04-25).