LIGAND BINDING BY RNA APTAMERS

RNA is unique among the three chief biological polymers. Like DNA, a simple mechanism exists for the replication of its primary structure. Like protein, it is capable of adopting stable tertiary folds. That combination has supported a long-standing conjecture that RNA is the first functional polymer in biology, and the RNA-life hypothesis argues that the first organisms used RNA to both propagate genetic information and to perform cellular functions.

In 1990, two laboratories independently capitalized on the dual function of RNA to evolve functional RNA molecules in the lab starting with large random RNA populations[1]. Larry Gold's lab coined the term SELEX (selective enrichment of ligands by exponential enrichment) to describe the process of identifying a small number of functional RNA molecules (called **aptamers** by the Szostak lab). The Gold lab prepared an RNA molecule capable of binding a protein, T4 DNA polymerase, with a dissociation contant of roughly 1×10^{-9} M, while the Szostak effort produced a series of RNA molecules capable of selectively binding a variety of dyes. From that starting point, some of the early work on aptamers focused on identifying the structural origins of selectivity and affinity of RNA aptamers for targeted ligands.

The SELEX Process

The key to generating a functional RNA is to start with a lot of RNA period. The hope is that within the haystack of many different RNA molecules, a few will possess the desired function. SELEX is designed to find those needles and to generate significant amounts of them. The process is defined in Table Ap.1 and Figure Ap.1.

As with all forms of evolution, selection should happen on a large sample of variants. Briefly one starts with a pool of synthetically generated single stranded DNA molecules, let's say 100 bases in length. The nature of chemical synthesis on a solid support permits the addition of mixtures of the four bases so that each growing strand will randomly have one added base. Typically, about a 50-base stretch will be random in sequence with the ends possessing defined sequence to allow replication. Ten micrograms of a 100-base oligomer will contain about 3×10^{-10} moles of DNA, or about 2×10^{14} molecules. There are 4^{50} different random sequences (1×10^{30}), so we would be sampling a miniscule fraction of all possible sequences. The DNA can be converted to RNA by the action of T7 RNA polymerase. So now the variation initially present as DNA is available as RNA.

Selection then acts upon the pool of RNA variants. Commonly, a chosen ligand is immobilized on a solid support and the pool of random RNA is given an opportunity to bind to the ligand, thus becoming immobilized to the solid support as well. Non-bound RNA can be washed away from the solid support, and then the bound RNA can be eluted by a variety of techniques (pH, salt, or an excess of free ligand).

[1] (a) A. Ellington & J. W. Szostak (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822. (b) C. Tuerk & L. Gold (1990) Systematic evolution of ligands by exoponential enrichment: RNA ligands to T4 DNA Polymerase. *Science* **249**, 505-510.

Table Ap.1. Steps in SELEX

1. Generate large number (~10^{15}) molecules of DNA possessing stretch of random sequence.
2. Transcribe DNA to RNA.
3. Test ability of RNA molecules to bind a ligand.
4. Remove failures, recover RNA molecules that succeed in passing test.
5. Convert successful RNAs to DNA.
6. Make multiple copies of DNA via PCR.
7. Repeat steps 2-6 as needed.

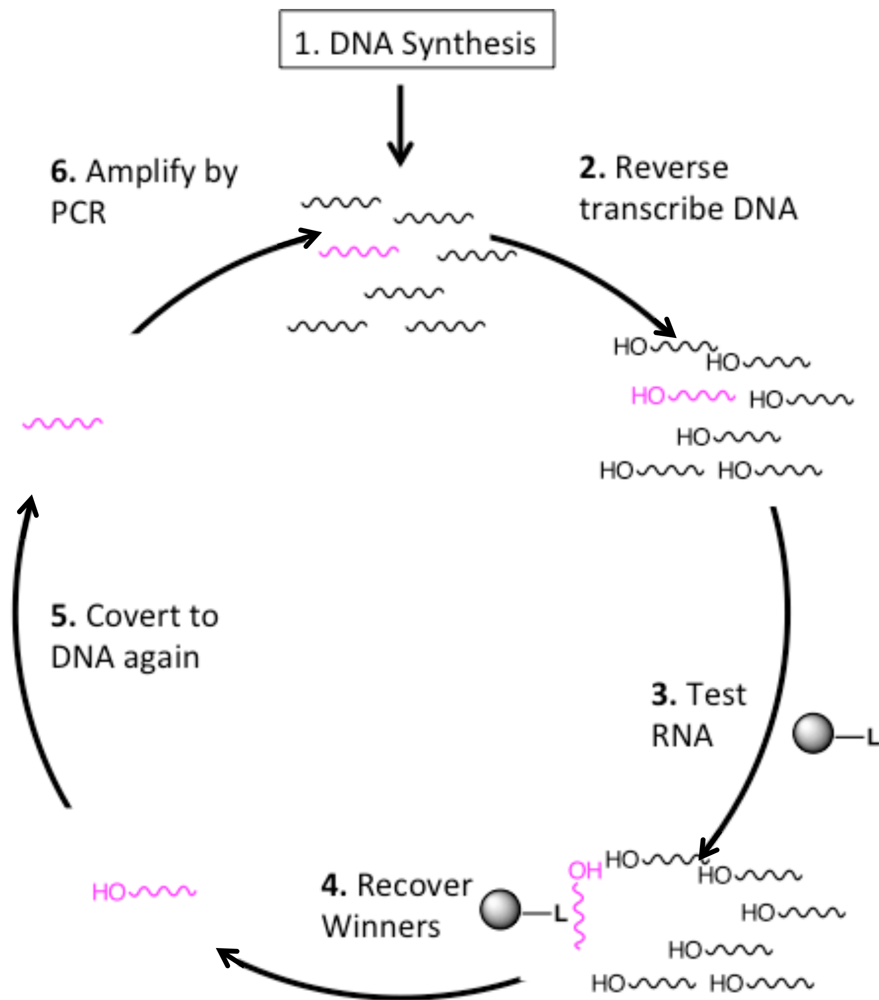

Figure Ap.1. Illustration of steps involved in SELEX. Note that step 6 does not reproduce failures, but rather will be a collection of winners. Some better, some worse. The next cycle will improve the quality of the winners.

The successful RNA molecules are then "reverse transcribed" into DNA again by the action of an enzyme called reverse transcriptase, and they are allowed to multiply through the use of the polymerase chain reaction which permits replication of a given DNA sequence into many, many copies. This creates a new pool of variants, albeit with less sequence diversity, for a second round of testing. They are converted again to RNA, tested with the immobilized ligand, and successful variants are retained. In these subsequent rounds, the competition between RNA molecules for the ligand becomes more intense, so fitter oligonucleotides are battling it out, providing for ever improved function.

In the end, usually after 5-7 cycles of selection, the successful RNA molecules are reverse transcribed back to DNA, they are then isolated from one another and pure "clones" of a single successful RNA molecule are isolated. Even after this many cycles, there is typically a diverse population of winners. The name "aptamer" takes meaning now. It is a collection of molecules that share a function (an aptitude) if not a structure. The goal now becomes determining the origin of the function in the context of the structure.

A Theophylline-Specific RNA Aptamer

Theophylline is a modified purine that can be used as a bronchodilator in the treatment of asthma. It is closely related to the recreational drug, caffeine (Figure Ap.2). Arguably, a selective means for detecting theophylline against a high background of caffeine could have useful diagnostic value in the case of a potential toxic dose of theophylline. Stepping away from the particular value of such a tool for detection, it would clearly require a receptor with strong selectivity for one of two closely related compounds.

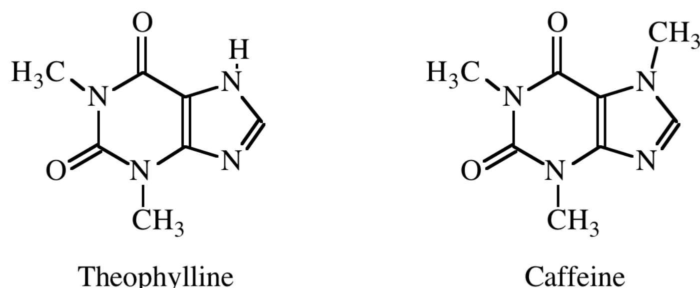

Figure Ap.2. Structures of theophylline and caffeine.

An early study in aptamer preparation tackled this problem through a combination of positive and negative selective pressure.[2] Starting with a pool of 10^{14} molecules of RNA containing a stretch of 40 random nucleotides (4^{40} possible structures, or about 10^{24}), selection for a theophylline binding receptor was achieved by running the pool of RNA over a column containing theophylline-modified agarose, where the covalent attachment was to the N1 methyl group. The non-binding RNA was eluted and bound RNA was recovered by elution with free theophylline.

[2] Jenison et al. (1994) "High-Resolution Molecular Discrimination by RNA" *Science* **263,** 1425.

In the first round, approximately 0.05% of the total RNA in the pool bound to the column. After seven rounds of selection and amplification of the RNA eluted by theophylline, 62% of the RNA would bind. At this point, positive selection had likely generated a pool of high-affinity theophylline binding RNA. But not specific theophylline-binding RNA. Many members of that pool would likely bind caffeine as well.

To weed out the non-specific binders, a negative selection step was introduced. After the 7th round, the amplified RNA pool was exposed to the theophylline column and then eluted with caffeine. In the first pass at this negative selection step, 99.7% of the RNA eluted. The remaining RNA was then eluted successfully with theophylline and amplified. In the third round of negative selection with caffeine, only 20% eluted.At that point, several of the RNA winners were converted to DNA, cloned and sequenced. In the process, it became clear that a common sequence was visible in the various individual survivors.

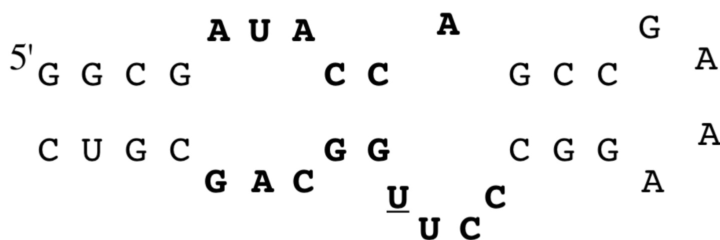

Figure Ap.3. Secondary structure of conserved sequence elements in theophylline binding aptamers.

Against unrelated background sequence, the two segments 5' AUACCA and 5' CCUUGG(C or A)AG (separated by >8 bases) were found. The AUACCA could come in front of the $CCUUGG^{C}/_{A}AG$ sequence or after. All possessed the ability to bind theophylline with a K_d of between 0.5 and 3 μM. Analysis of the sequences suggested the secondary structure presented in Figure Ap.3.

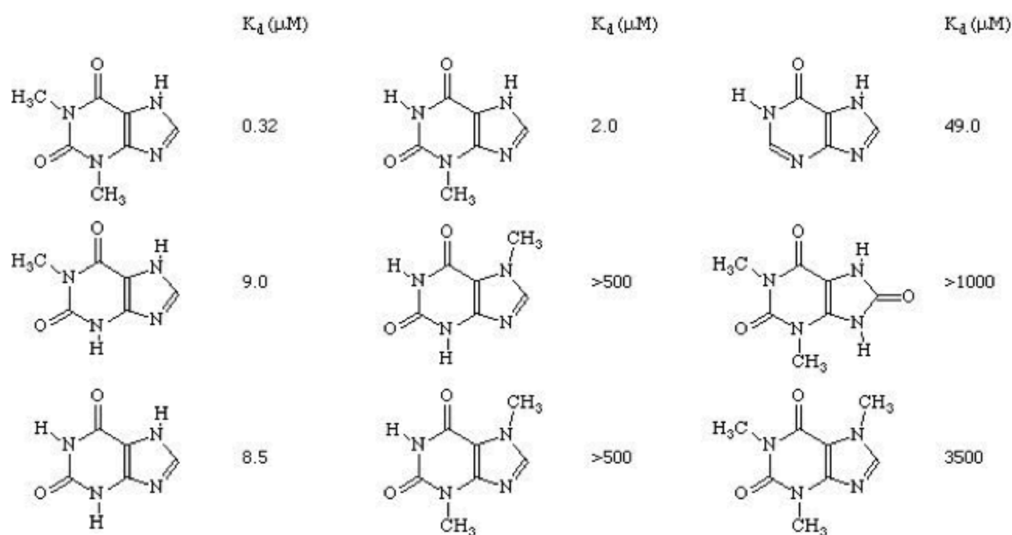

Figure Ap.4. Dissociation constants for several theophylline analogs.

Importantly, the aptamers generally showed selectivity for theophylline over caffeine. One candidate was selected for functional analysis (Figure Ap.4). While there was only modest selectivity for the methyl group of N1, not surprising given that the linker to the column was placed there, strong selectivity was observed for modifications elsewhere in the purine ring. In particular, at N7, where caffeine has a methyl group, binding was decreased 10^4-fold when that methyl group was added. A preliminary model predicted that steric exclusion would impede binding a purine with a methyl group at that position, and H-bonds to the C2 and C4 carbonyls would help enhance affinity, while a hydrophobic binding pocked for the methyl group at N3 would also provide selectivity.

In a subsequent NMR structure of a theophylline-specific aptamer, the basis for affinity and specificity was made clear (Figure Ap.5).[3] General non-specific base stacking above and below the bound ligand provide a good binding pocket for the purine ligand, but specificity is achieved within the plane of the molecule. C2 and U4 of the C$\underline{C}$U$\underline{U}$GG$^{C}/_{A}$AG form hydrogen bonds with substituents on the theophylline ring. In particular, H-bond donation for C2 to N7 of theophylline clearly excludes a methyl group at that position (Figure Ap.5). Interestingly, there is no clearly specific interaction with the methyl group at N1 or the carbonyl at C2.

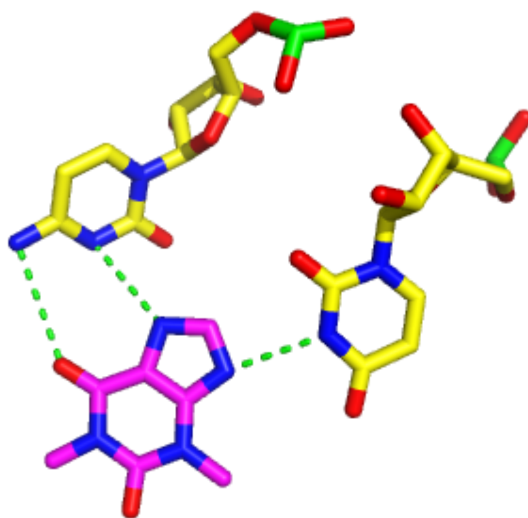

Figure Ap.5. Specific hydrogen bonding interactions between aptamer bases (yellow carbons) and theophylline (magenta carbons). The position of cytosine 2 (above) clearly prevents the N7 methyl substituent in caffeine.

---

[3] Zimmerman et al. (1997) "Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA." *Nature Struct. Biol.* **4**, 644.