J Mol Evol (2015) 81:150–161
DOI 10.1007/s00239-015-9708-6

ORIGINAL ARTICLE

Analysis of In Vitro Aptamer Selection Parameters

Maureen McKeague[1] · Erin M. McConnell[1] · Jose Cruz-Toledo[2] · Elyse D. Bernard[3] · Amanda Pach[1] · Emily Mastronardi[1] · Xueru Zhang[1] · Michael Beking[1] · Tariq Francis[1] · Amanda Giamberardino[1] · Ashley Cabecinha[1] · Annamaria Ruscito[1] · Rocio Aranda-Rodriguez[3] · Michel Dumontier[2,4] · Maria C. DeRosa[1]

Received: 9 September 2015 / Accepted: 22 October 2015 / Published online: 3 November 2015

Abstract Nucleic acid aptamers are novel molecular recognition tools that offer many advantages compared to their antibody and peptide-based counterparts. However, challenges associated with in vitro selection, characterization, and validation have limited their wide-spread use in the fields of diagnostics and therapeutics. Here, we extracted detailed information about aptamer selection experiments housed in the Aptamer Base, spanning over two decades, to perform the first parameter analysis of conditions used to identify and isolate aptamers de novo. We used information from 492 published SELEX experiments and studied the relationships between the nucleic acid library, target choice, selection methods, experimental conditions, and the affinity of the resulting aptamer candidates. Our findings highlight that the choice of target and selection template made the largest and most significant impact on the success of a de novo aptamer selection. Our results further emphasize the need for improved documentation and more thorough experimentation of SELEX criteria to determine their correlation with SELEX success.

Keywords Aptamer · In vitro selection · SELEX · Aptamer target · Database

Authors Maureen McKeague and Erin M. McConnell contributed equally.

Electronic supplementary material The online version of this article (doi:10.1007/s00239-015-9708-6) contains supplementary material, which is available to authorized users.

✉ Michel Dumontier
michel.dumontier@stanford.edu

✉ Maria C. DeRosa
maria.derosa@carleton.ca

[1] Chemistry Department, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada

[2] Biology Department, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada

[3] Environmental Health Science and Research Bureau, Health Canada, Tunney's Pasture, Ottawa, ON K1A 0K9, Canada

[4] Stanford Center for Biomedical Informatics Research, Department of Medicine, Stanford University, Stanford, CA 94305, USA

Introduction

Traditionally, nucleic acids were only associated with information storage and genetic coding whereas proteins were classified as the functional workhorses of the cell (Deigan and Ferre-D'Amare 2011; Gilbert 1986; Waldrop 1989). However, the last few decades have revealed that nucleic acids possess qualities and functions that are much more like proteins than previously thought. Indeed, nucleic acids able to form intricate tertiary structures that perform a variety of functions such as ligand binding, catalysis, and gene regulation have been discovered or engineered (Serganov and Patel 2007). In particular, one class of functional nucleic acids, known as aptamers, has emerged as novel and flexible molecular recognition tools due to their unique ability to recognize and bind to a specific target of interest (McKeague and Derosa 2012). Examples of these functional nucleic acids were discovered by three separate groups (Ellington and Szostak 1990; Robertson and Joyce 1990; Tuerk and Gold 1990) over two decades ago and have been an important ongoing research area in molecular biology, nucleic acid chemistry, and synthetic biology. The application of aptamers spans many areas including diagnostics, therapeutics, and environmental sensing (Famulok et al. 2007; McKeague et al. 2011; Nimjee et al. 2005) as

they offer significant improvement over protein-based molecular recognition technologies. These benefits have been substantially reviewed elsewhere (Jayasena 1999); briefly, aptamers are attractive for a multitude of applications due to their ease of chemical modification, stability, low molecular weight, and their rapid and reliable synthesis and production.

Natural aptamers that exist as the recognition component of genetic control elements known as riboswitches have been identified across all domains of life and bind to several metabolites (Barrick and Breaker 2007); however, synthetic aptamers that bind to a wide variety of targets can be generated de novo. These synthetic nucleic acids fold into complex structures and interact with their cognate target through a combination of hydrogen bonding, hydrophobic interactions, electrostatic interactions, and pi–pi stacking. Due to the complex and diverse nature of these interactions, there are currently very limited computational methods and rational design strategies that can be applied to this type of sampling problem. As a result, aptamers are almost exclusively identified through an in vitro selection procedure termed Systematic Evolution of Ligands through Exponential enrichment (SELEX). This procedure has allowed for the successful generation of aptamers that bind to a variety of targets including cells, proteins, small molecules, and ions. Many chapter protocols and reviews have been published detailing in vitro selection methods (Cho et al. 2009; Gold et al. 2012; Hall et al. 2009; Jhaveri and Ellington 2002; Sefah et al. 2010; Zhu et al. 2012), reporting on typical experimental conditions and strategies that have been successfully applied over the past two decades for a set of targets of interest. Furthermore, a handful of groups have described mathematical quantification of selection experiments and developed procedural guidelines to improve the selection success rate (Ciesiolka et al. 1996; Gold et al. 2010).

Despite the documented procedures and the expanding potential for aptamers, these ligands have not reached their full potential—antibodies remain the gold standard in molecular recognition and few examples of clinical success can be named for aptamer-based therapeutics (Baird 2010; Mattice and DeRosa 2015). There are two main bottlenecks that currently limit the wide-spread application of these nucleic acid ligands compared to their protein counterparts. First, there is no standardized method for characterizing aptamer functionality, and no repository to search this validation data exists. As an example, antibodies are among the most frequently used tools in basic science research and in clinical assays, thus vendors perform rigorous validation to determine the selectivity and reproducibility in the context for which they are to be used (Bordeaux et al. 2010). In contrast, applying a reported aptamer to a new application is challenging and requires significant optimization due to a general lack of validation standards in the field (McKeague et al. 2015). As another example, numerous protein sequence and structure databases exist that are maintained through international consortia or government agencies to ensure they can be easily accessed, managed, and updated (Apweiler et al. 2004). Fortunately, the past decades have seen the development of several individually maintained aptamer databases that serve as a repository for reported aptamers. Examples include the Aptamer Database (Lee et al. 2004), SELEX-DB (Ponomarenko et al. 2002), HTP-SELEX (Jagannathan et al. 2006), and RiboAptDB (Thodima et al. 2006). Most recently, Aptamer Base (Cruz-Toledo et al. 2012) was developed that additionally maintains detailed properties about the aptamers and the conditions in which they were selected. The open-access format of this database, available through Wikidata, allows original aptamer data to be enhanced by the manual addition of further information. While currently not applied for this purpose, the flexible format of this newer database may serve as a useful tool to provide centralized, validated information about aptamers in the future.

The second major bottleneck limiting the wide acceptance of aptamers is that the in vitro selection method used to identify new aptamers is time consuming, generally not high-throughput, and often results in failure (Gold et al. 2010; Mascini et al. 2012). This challenge can again be compared to research problems associated with their protein counterpart ligands. For example, crystallization is the most serious bottleneck in high-throughput protein-structure determination by diffraction methods (Price et al. 2009). Similar to in vitro aptamer selection experiments, the current approach to protein crystallization is empirical. Due to the large number of variables involved (e.g., concentration, temperature, pH) little guidance is available to systematically identify solution conditions that promote crystallization. As a result, proteins are screened against arrays of many chemical conditions that have successfully yielded crystals in the past (McPherson and Cudney 2014). However, unlike the aptamer field, significant work dedicated to understanding the physical properties and parameters that control protein crystallization has been performed (Carter and Yin 1994; Fusco et al. 2014; Price et al. 2009).

To mitigate the abovementioned challenges in aptamer research, we leveraged the information provided in Aptamer Base to perform an initial analysis of trends used in aptamer selection over 23 years. While there have been a few initial systematic studies to improve certain selection strategies (Carothers et al. 2010; McKeague and DeRosa 2014; Velez et al. 2012), a more thorough analysis is required to make conclusions applicable to the field as a whole. Through our analysis, we anticipate we will gain insight about the parameters and conditions that result in successful aptamer selection.

Materials and Methods

Data

All data used in these analyses were obtained from the free, openly licensed community-built Resource, Aptamer Base (http://www.freebase.com/base/aptamer) which will soon be imported to Wikidata (Cruz-Toledo et al. 2012). The Aptamer Base includes descriptions about 492 SELEX experiments published between 1990 and 2013. This dataset includes 2334 unique aptamer sequences to 569 unique targets, where each aptamer-target interaction has an associated dissociation constant. Importantly, each individual SELEX experiment contains information about the employed nucleic acid library composition and length, the experimental conditions (e.g., buffer, pH, temperature, cation concentration), the number of selection rounds, and the selection methods employed.

Data Retrieval

Data from the Aptamer Base were programmatically accessed through Freebase's application programming interface (API) using the Metaweb Query Language. A suite of Java 1.6 programs were developed which implement Google's latest Java API clients (http://code.google.com/p/google-api-java-client/) to query the Aptamer Base. For examples of individual queries see (Cruz-Toledo et al. 2012) supplemental information.

Data Analysis

Each SELEX experiment may yield one or many aptamer sequences possessing a wide range of affinities (reported as a dissociation constant, K_D) for the selected target. Typically, the final library will contain sequences with a range of affinities for the target. The aptamer sequence with the lowest K_D (highest affinity) is considered the aptamer of interest. The Aptamer Base has included all sequences from each SELEX experiment that were characterized and have a reported K_D. However, for all of our analyses, we have only used the lowest K_D for each individual SELEX experiment.

Statistical Analysis

For each analysis, equality of variance was evaluated by either Levene's or Bartlett's test. Where appropriate, either the parametric One-way ANOVA or non-parametric Kruskal–Wallis Rank-Sum test was used to compare the mean Log10(K_D) or mean rank of the Log10(K_D), respectively. When a significant difference between samples was revealed, either Tukey's HSD (honest significant difference) or the Mann–Whitney U test were employed to analyze the differences between the groups. Correlations between the Log10(K_D) and the variable of interest were examined by linear regression.

Results and Discussion

Using the detailed information present in Aptamer Base, we sought to examine the relationships between aptamer affinity and user-defined parameters that can be manipulated within an in vitro selection experiment. Specifically, we examined the trends in selections reported in the literature throughout the past 23 years, with a focus on library design, target of interest, and the experimental methods and conditions. For the analysis, we compared each of the selection parameters using the highest aptamer affinity per experiment (reported as the dissociation constant, K_D) as our measure of fitness score (i.e., the higher the affinity of any resulting aptamer, the more successful the selection). The Log10(K_D) was used for plotting and easy visual comparison.

Library Design

Nucleic Acid Type

The majority of early in vitro experiments sought to select RNA aptamers as it was generally considered that only RNA could fold into functional motifs (Gilbert 1986). However, a few years following the initial aptamer selection experiments, several groups reported DNA aptamers that were capable of binding molecular targets with high affinity, indicating that, single-stranded DNA exhibits a propensity for forming intricate tertiary structures comparable to that of RNA (Ellington and Szostak 1992; Walsh and DeRosa 2009). Additional motivation for the selection and implementation of DNA-based aptamers stems from the lack of the 2′ hydroxyl of the DNA sugar. This leads to improved chemical and biological stability compared to RNA, a useful characteristic for many applications of aptamers (Walsh and DeRosa 2009). Despite this initial finding, for many years researchers continued to preferentially select RNA aptamers due to the suggestion that RNA results in higher affinity binders. Figure 1a highlights the frequency of DNA or RNA library templates in reported aptamer selection experiments over 23 years.

Over the entire time period from 1990 to 2013, there was a slight bias observed for RNA over DNA (Fig. 1b). Interestingly, there has been a shift in this bias toward DNA-based selection beginning in the publication year 2008. When the frequency of DNA versus RNA selections

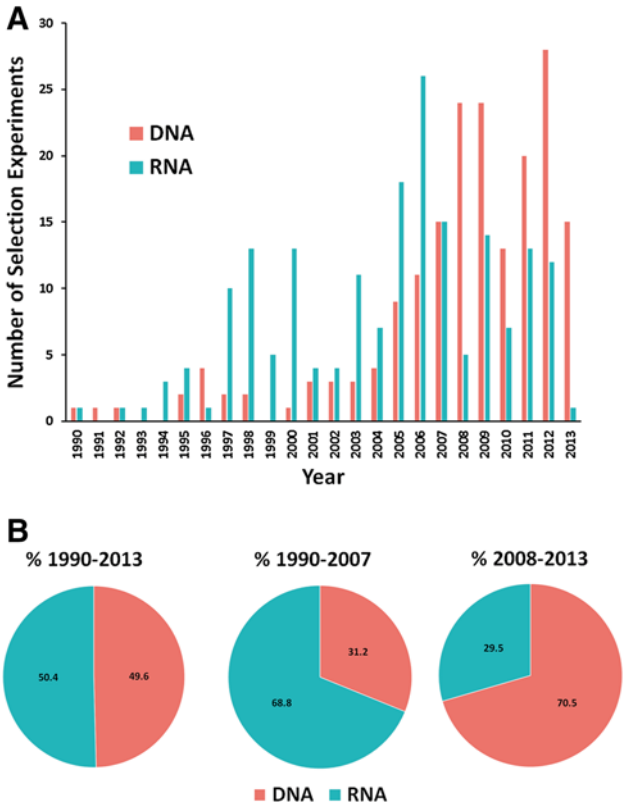

Fig. 1 **a** Frequency of selection experiments utilizing DNA (*pink*) or RNA (*blue*) nucleic acid templates from 1990 to 2013. **b** Distribution of SELEX experiments with DNA (*pink*) or RNA (*blue*) templates over time (Color figure online)

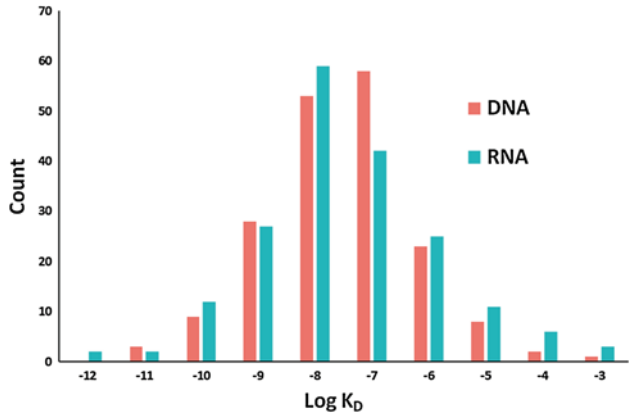

Fig. 2 Distribution of the highest measured aptamer affinity for each experiment (reported as the dissociation constant, Log10(K_D)) for DNA—(*pink*) and RNA—(*blue*)-based selection experiments (Color figure online)

was analyzed as subpopulations of 1990–2007 and 2008–2013, a dramatic change in % occurrence was observed. In the years up to 2007, ~70 % of selection experiments were based on RNA templates. During the period of 2008–2013, use of DNA templates increased dramatically to ~70 % of selection experiments (Fig. 1b).

Using the information in the database, we sought to investigate whether either aptamer type presented a significant binding affinity advantage. The histograms representing the dissociation constants for DNA ($n = 185$) and RNA ($n = 189$) aptamers reveal there is no significant difference between the mean Log10(K_D) for the different aptamer types (t test; $p = 0.69$; see Fig. 2), suggesting neither DNA nor RNA yield aptamers with better affinity. Experimental evidence in the literature supports the finding that neither DNA nor RNA aptamers have systematically higher affinity for their targets. For example, DNA and RNA aptamers for several targets have been identified with similar binding affinities (e.g., tetracycline, ATP, dopamine, kanamycin) (McKeague and Derosa 2012). Taken together, the absence or presence of the 2′OH in the sugar backbone, or of the methyl group on pyrimidine bases has no differential binding effect. As modified nucleic acids (within the sugar phosphate backbone, or modifications to the base) are included more frequently in new aptamer selections (Tolle et al. 2015), it will be useful and interesting to statistically re-examine the potential effect on aptamer affinity. We also examined any potential significance in the actual library composition (percentage of each nucleobase), results are discussed in the Supplemental Material. While percent GC content, as well as percent G content, was found to be elevated in aptamers in general, no statistically significant correlation could be found between base content and aptamer affinity (see Section S1 and Figs. S1–S8).

Template Length

The choice of random region length (length of N) for a selection experiment is typically made in a relatively arbitrary fashion (Velez et al. 2012). Random regions as short as 22 nucleotides (Lozupone et al. 2003) and as long as 200 (Li et al. 1996) have successfully resulted in aptamer discovery, however, random regions of N40 to N70 are more commonly employed. This smaller range of random region length choice serves to strike a balance between (1) ensuring appropriate coverage of sequence space, where shorter random regions allow greater coverage (Silverman 2009) and (2) allowing for greater structural complexity, where longer random regions allow greater complexity of secondary and tertiary structure (Sabeti et al. 1997). This large, random distribution of choice in library length is similarly observed for both RNA and DNA selections over the last two decades (see Fig. 3).

To date only a few groups have explored the importance of random region length for in vitro selection experiments. For example, (Legiewicz et al. 2005) examined random regions from length 16–90 to identify RNA aptamers to

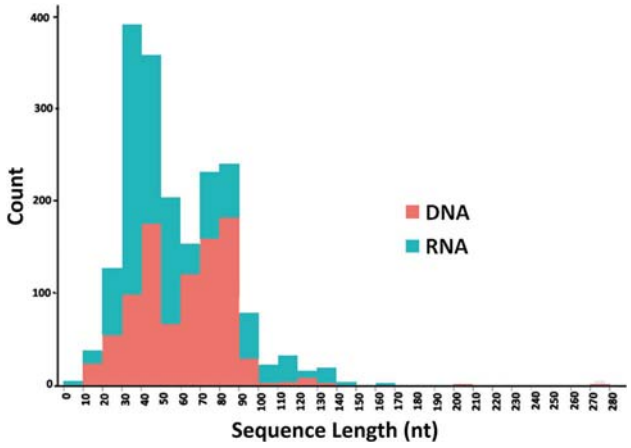

Fig. 3 Overlay of the distribution of the library length (random region length) of each selection experiment for DNA—(*pink*) and RNA—(*blue*) (Color figure online)

isoleucine and determined that a random region of length 50 or 70 was optimal. This has not been verified for other aptamer selections, thus researchers cannot be confident in choosing the best compromise between sequence space coverage and structural complexity. Therefore, we sought to determine whether there was an observable effect of length on aptamer binding. Our initial linear regression analysis comparing the template length and resulting aptamer affinity reveals no significant correlation ($R = 0.01$) between template length and K_D (see Supplemental Fig. S9). More recently, (Velez et al. 2012) established that different lengths may be better suited for particular functions. To clarify this result, we further examined the correlation between template length and aptamer dissociation constant by specific aptamer type and target type (see Table 1).

From these data, we can conclude there is no significant correlation between nucleic acid type and template length on aptamer affinity ($R < 0.5$). However, from the correlation coefficient (R) there is a slightly stronger positive correlation for DNA ($R = 0.26$) than for RNA ($R = 0.19$). When examining the effect of different aptamer targets on the relationship of template length compared to K_D, some significant correlations were observed. For example, there is a positive correlation between the K_D and template length for peptide targets ($R = 0.50$) and a negative correlation for carbohydrate targets ($R = 0.50$). For viruses, there is a weaker positive correlation. Finally, the positive correlations can be interpreted to suggest that shorter template lengths may result in improved aptamer affinity in all instances (nucleic acid type and target type) with the exception of carbohydrate targets, which would require longer templates for higher affinity. It is not apparent why this might be the case. Further experimental work may reveal the effect of specific functional group, structure, and molecule weight on aptamer selections. As larger datasets become available, significant correlations for additional target types or a direct effect of nucleic acid type on the relationship between K_D and template length may be revealed. However, we cannot rule out the possibility that variable length on its own has very minimal impact on aptamer affinity; specific sequence and structure are more important (Luo et al. 2010). As an immediate resolution to this selection challenge, one strategy would be to employ libraries of several variable lengths; while laborious, this method has improved the success of some selections (Velez et al. 2012).

Aptamer Target

We sought to investigate the effect of particular target type on the ability to select high-affinity aptamers. While the majority of aptamers are selected against proteins, there are a variety of other target types for which aptamers have been successfully generated (see Fig. 4).

Our motivation stemmed from the earlier observation that a particular target type revealed a stronger correlation between library length and dissociation constant (for example, peptide and carbohydrate targets; Table 1). Furthermore, excellent work has been performed (Carothers et al. 2010) in an attempt to determine the effect of target structure and size with binding affinity for small molecules. Specifically, Carothers determined that the target molecular weight was proportional to the resulting aptamer affinity (larger targets resulted in lower K_D values). As demonstrated in Fig. 4b, the variety of targets for which aptamers have been selected has expanded to include viruses and whole cells in recent years.

Table 1 Correlation between template length and aptamer dissociation constant (K_D) by nucleic acid type and target type

	Slope of linear regression trendline	r^2	R	Significant at 0.5
Nucleic acid type				
RNA	0.012	0.0345	0.19	No
DNA	0.0081	0.0684	0.26	No
Target type				
Peptide	0.041	0.252	0.50	Yes
Virus	0.028	0.188	0.43	No
Carbohydrate	−0.053	0.248	0.50	Yes

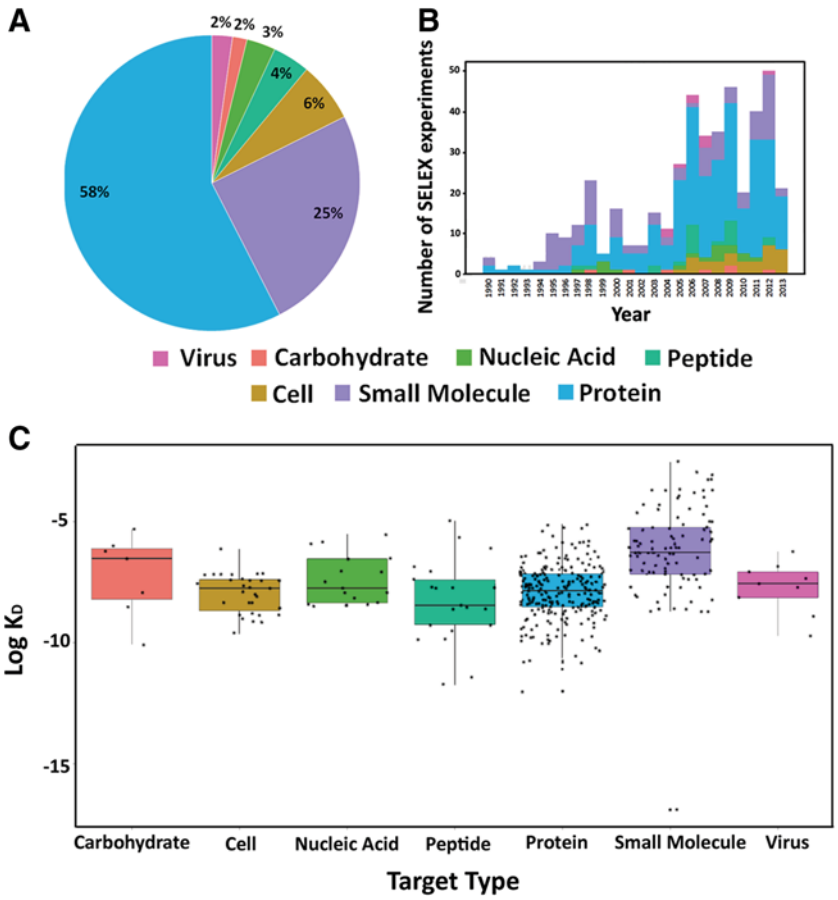

Fig. 4 **a** Percentage of aptamers grouped by target type, reported in the literature from 1990 to 2013. **b** Occurrence of selections by target type over the same time period. Selection data for each year are stacked. **c** Box plot of target type versus aptamer affinity, reported as the dissociation constant, Log10(K_D) (Color figure online)

Thus, we compared the affinity of resulting aptamers for the broad seven encompassing groups (target types) of all aptamer targets. Figure 4c highlights the mean and standard deviation of the K_D values for each target group. An initial comparison revealed there is a significant difference between the affinity of the mean Log10(K_D) values of at least two target types (Kruskal–Wallis: $p \ll 0.05$). To further confirm this observation, we performed a post hoc analysis using a pairwise Mann–Whitney U test to compare the mean rank of the Log10(K_D) of each of the target groups. Consistent with observations by Carothers et al. (2010), aptamers that bind to small molecule targets may have lower affinity compared to all target groups. The results are summarized in the Table S1. In each case, with the exception of viruses, this post hoc analysis revealed that small molecule targets result in lower-affinity aptamers. It is important to note there was a large discrepancy in sample sizes and more data will be required to establish definitive trends.

Selection Conditions

As in any chemical experiment, the experimental conditions (i.e., selections conditions) must be carefully considered. Selection condition data from the Aptamer Base were examined to determine if a correlation existed between specific selection conditions such as metal cation concentration, buffering agent of the selection buffer, pH, and temperature compared to the measured affinity of aptamers. Where possible, we compared our large-scale Aptamer Base observations to aptamer-specific laboratory experiments and findings. These comparisons typically draw from a handful of aptamers that are frequently used in proof-of-concept studies and are thus well characterized. These specific examples highlighting the effects of altering selection conditions on the K_D will be discussed. The combination of our bioinformatics analysis with aptamer-specific laboratory experiments can help us make meaningful conclusions about aptamer selection conditions.

Metal Cation Concentration

The presence and concentration of metal cations can significantly affect the secondary structure, the stability, and the affinity of selected aptamers. In general, increasing the concentration of metal cations can lead to shielding of the negative charge of the DNA backbone, effectively eliminating potential electrostatic binding interactions of the

DNA to the target (Hianik et al. 2007). Likewise, the metal cation concentration may impact the charge of the target, or cause it to aggregate (in the case of some proteins) leading to decreased aptamer-target affinity. Additionally, variations in salt concentration can lead to either ion- or concentration-dependent secondary structural transitions. In these cases, one conformation may bind the target, but affinity may be significantly diminished or eliminated following structural transition. In fact, some sensor designs exploit this property of aptamers (Catherine et al. 2014).

To determine if metal cation concentration used during selection affected the affinity of aptamers, the relationship between Mg^{2+}, Na^{+}, and K^{+} and the K_D were independently investigated. Briefly, to examine the relationship between metal cation concentration and reported K_D for nucleic acid type and target type, the Log10(K_D) was plotted against the Log10(cation concentration). In almost all cases, the observed correlations were positive, suggesting lower K_Ds (higher affinity) are obtained with lower metal cation concentrations. The exception was negative correlation between virus target type and Na^{+} concentration.

Our observations are consistent with data reported for several aptamers commonly used in proof-of-concept studies. For example, (Hianik et al. 2007) showed that the sensitivity of their human α-thrombin aptamer-based sensor decreased with increasing Na^{+} concentration. As a second example, the ATP-binding aptamer was evaluated in terms of target retention to a column. In this study, analyte retention decreased as Na^{+} concentration was increased and retention increased with increasing Mg^{2+} concentration (Deng et al. 2001). Finally, the cocaine aptamer, frequently used in structure switching or bipartite aptamer studies, actually showed the highest affinity in the absence of Na^{+} with a marked decrease in affinity in the presence of Na^{+} (Neves et al. 2010).

Preliminary analysis (summarized in Table 2) revealed there was no significant correlation between metal cation concentration and the apparent K_D values of different nucleic acid types. Interestingly though, some weak positive correlations were observed. Overall there was a stronger correlation between metal cation concentrations versus K_D for RNA than what was observed for DNA. Comparing between cations within the RNA group, there was a stronger correlation between Na^{+} concentration and K_D than either Mg^{2+} or K^{+} concentration and K_D. There was a weak correlation between Mg^{2+} concentration and K_D for DNA, but Na^{+} and K^{+} showed almost no correlation to K_D.

The ability of an aptamer to bind to a target is dependent on the intermolecular interactions between the target and the aptamer. Each target type has vastly different characteristics including charge, polarity, hydrophobicity, chemical composition, and size. Small-scale studies have begun to experimentally demonstrate a correlation between the strength of aptamer-target interactions (Carothers et al. 2010) and these different parameters. As a starting point, we investigated whether there was a correlation between metal cation concentration and the reported K_D for different subsets of target types where data were available (Table 2).

Overall, stronger correlations between metal cation concentration and K_D for target type versus nucleic acid type were observed. In a practical sense, this may mean that when deciding on the metal cation composition of the selection/affinity experiment buffers one should more carefully consider their desired target than the nucleic acid type. From these data, we can conclude that the correlation between metal cation concentration and K_D was not significant; however, compared to nucleic acid type, there were relatively strong correlations observed between each metal cation and K_D for small molecule targets. However, significant positive correlations were observed for virus and peptide target types in the presence of Mg^{2+}. For these targets, researchers may consider using lower concentrations of Mg^{2+} cations to improve the likelihood of isolating high-affinity aptamer candidates.

Buffering Agent

The choice of the buffering agent used in the selection buffer is based on user preference. However, there is literature precedence that the buffering agent may affect aptamer affinity. For example, (Deng et al. 2001) examined the effect of buffer on binding affinity of the ATP aptamer by determining retention and resolution of multiple targets from an affinity chromatography column. Interestingly, switching from tris(hydroxymethyl) aminomethane (Tris) to phosphate-buffered solutions led to better resolution between targets. To examine whether or not there was a significant difference between the buffering agent used in the selection buffer and the resultant measured affinity of the aptamers, the mean Log(K_D) of each buffering group was compared (Fig. 5 Kruskal–Wallis test, $p = 0.025$).

Since Tris-buffered saline was the most commonly reported buffering agent used (~50 % of the experiments sampled), the Mann–Whitney U test was used to compare mean ranks between Tris-buffered saline and less commonly used buffering agents. A significant difference between the mean rank of Tris-buffered saline and Phosphate-buffered saline was revealed ($p = 0.041$). The difference between Tris-buffered saline and HEPES was not significant. Approximately, 5 % of the SELEX experiments sampled did not properly describe the buffering agent used. These conclusions should be interpreted carefully as further analysis when more data on buffering agent become available may

Table 2 Correlation (represented by the correlation coefficient, R) and slope of the linear regression trendline (m) of cation concentration and K_D based on nucleic acid type and target type

	Mg^{2+}	Na^{+}	K^{+}
Nucleic acid type			
RNA	$R = 0.38$; $m = 1.5$	$R = 0.48$; $m = 2.8$	$R = 0.21$; $m = 0.34$
DNA	$R = 0.21$; $m = 0.66$	$R = 0.06$; $m = 0.25$	$R = 0.06$; $m = 0.12$
Target type			
Small molecule	$R = 0.43$; $m = 1.5$	$R = 0.46$; $m = 2.2$	$R = 0.39$; $m = 0.62$
Virus	$R = 0.88$*; $m = 2.5$	$R = 1.00$*; $m = -3.7$	(n/a)
Peptide	$R = 0.58$*; $m = 2.2$	(n/a)	(n/a)
Carbohydrate	(n/a)	$R = 0.38$; $m = 4.1$	(n/a)

* Significant correlation (significance at $R \geq 0.50$)

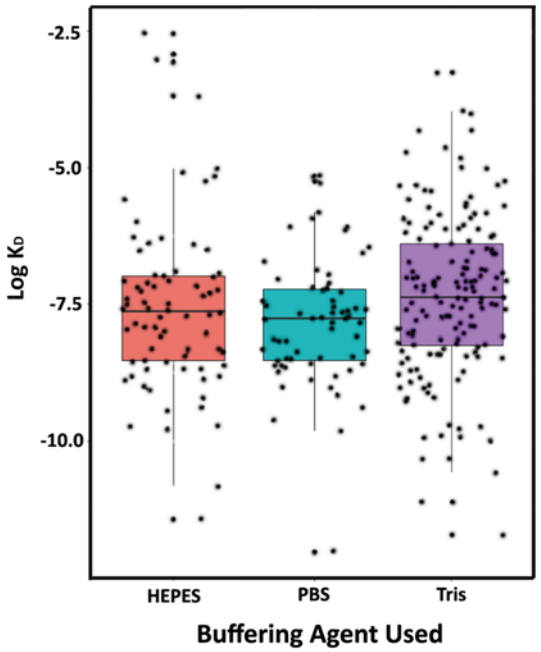

Fig. 5 Box plot of buffering agent of selection buffer versus aptamer affinity, reported as the dissociation constant, Log10(K_D)

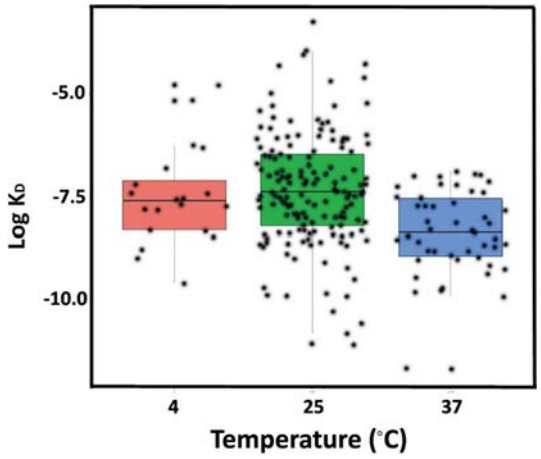

Fig. 6 Box plot of temperature used versus aptamer affinity (reported as the dissociation constant, Log(K_D))

reveal different trends. Furthermore, the high percentage of SELEX experiments lacking appropriate experimental detail serves as an important reminder to the community to provide sufficient information to repeat an in vitro selection experiment (Cho et al. 2009).

Temperature

In our dataset, selections were performed at either 4, 25, or 37 °C. The most common selection condition was at 25 °C followed by 37 °C, and 4 °C making up ~70, ~21, and ~9 % of selections respectively. The choice of selection temperature is typically dependent on the target and application. For example, selections done at 37 °C mimic physiological conditions for biologically relevant experiments whereas selections at 4 °C allow for the reduction of non-specific cellular uptake of oligonucleotides (Sefah et al. 2010). A comparison of the aptamer affinity as the Log10(K_D) values for each temperature group is represented as a box plot in Fig. 6.

We next examined the effect of temperature on the affinity of selected aptamers using a One-way ANOVA ($p \ll 0.05$). Post hoc analysis by Tukey's HSD revealed a significant difference between the mean Log10(K_D) of 25 and 37 °C ($p \ll 0.05$) as well as 4 and 37 °C ($p = 0.037$). In both cases, the affinity of aptamers selected at 37 °C was higher (lower K_D values); however, there was no significant difference between 4 and 25 °C ($p = 0.831$). In addition to this finding, many of the promising applications for aptamers are for use in biological or in vivo samples; thus, where possible, aptamer selections at 37 °C may be a useful experimental strategy; future experimental work should investigate this further.

pH of the Selection Buffer

pH is known to influence structure, stability, and charge/protonation of both the nucleic acids and targets, and thus

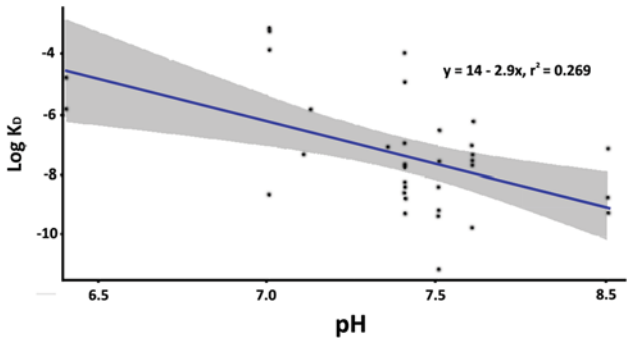

Fig. 7 pH of selection buffer versus aptamer affinity (reported as the dissociation constant, Log(K_D)

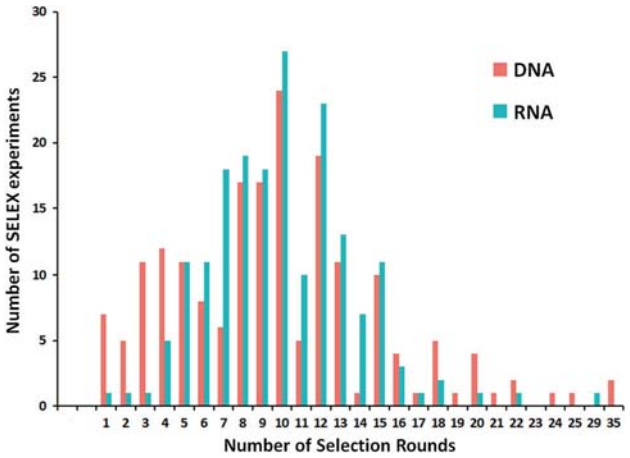

Fig. 8 Frequency of SELEX experiments by number of selection rounds for DNA (*pink*) and RNA (*blue*) aptamers (Color figure online)

significantly alter the aptamer-target binding interaction. For example, the affinities of the thrombin, ATP, and cocaine aptamers were all shown to be dependent on solution pH (Deng et al. 2001; Hianik et al. 2007; Neves et al. 2010).

Here, we sought to investigate the correlation of the initial selection buffer pH on aptamer affinity. Log10(K_D) was plotted against the pH of the selection buffer (Fig. 7). Our data showed a weak statistically significant negative correlation ($R = 0.51$) between Log10(K_D) and pH of the selection buffer. This might suggest that aptamers isolated from an experiment employing a buffer with higher pH have improved affinity to their target. Despite this observation, it is important to consider the practical impact of altering the pH in an experiment. For example, the selection buffer should be chosen to be consistent with the desired application.

Selection Methods

Over the past two decades, many variations of SELEX have been reported in an attempt to either improve the success rate of SELEX, to generate aptamers with novel designs and functions, or to increase the process throughput.

Number of Selection Rounds

Typically, SELEX requires approximately 5–15 rounds (Silverman 2009). This was confirmed by plotting the frequency of SELEX experiments against the number of selection rounds (Fig. 8). The mean number of selection rounds was determined to be 10 ± 6 and 10 ± 4 for DNA and RNA, respectively.

The number of selection rounds performed depends on several factors. Typically, researchers monitor enrichment as an indication of selection completion. Reduced number of rounds will decrease experimental cost and improve throughput, but more rounds may be needed to achieve sufficient enrichment.

To examine whether the number of selection rounds affected the affinity of selected aptamers, the Log10(K_D) was plotted against the number of selection rounds. No correlation between number of selection rounds and Log10(K_D) was observed (Fig. S10). This is not surprising given that in a typical SELEX experiment, each round serves to only reduce the diversity of the population and there is little to no opportunity for mutation or shuffling to evolve the library. Therefore, with each additional round, the number of potential aptamer candidates remains the same. However, not doing enough rounds may result in a high diversity of the final population. In the past, low-throughput sequencing was used to obtain aptamer candidates; therefore, with a high population diversity, more sequencing and screening of a larger number of candidates may be required, thus reducing throughput and increasing associated costs (McKeague et al. 2015). Recently, high-throughput sequencing (HTS) has greatly improved this potential selection challenge. Particularly, researchers have found that reaching a threshold enrichment as an experimental end-point is less important. Open-access software interfaces to process the large amount of data have been developed and should facilitate the broader use of HTS to aid in future selections (Alam et al. 2015; Hoinka et al. 2014).

SELEX Methodology

Over the past two decades, significant changes have been made to the selection process, including changes to the "SELEX Method," "partitioning method," and "recovery method" (see Supplementary Material for a full description of the terms used for our analysis). Here we refer to the

SELEX method specifically as an adaptation of the original in vitro selection methods first described by (Ellington and Szostak 1990; Robertson and Joyce 1990; Tuerk and Gold 1990) that improves the affinity of selected aptamers, allows for the selection of unconventional targets, or streamlines the selection procedure. We define partitioning method as the method by which the oligonucleotides with affinity for the target are separated from oligonucleotides with little or no affinity for the target. Finally, recovery method is defined as a physical or chemical method used to separate binding oligonucleotides from the target prior to PCR amplification. The goal of changing these methods is typically to improve the partitioning coefficient or efficiency, to incorporate improved methods for monitoring enrichment, or to improve the throughput (Carothers et al. 2006; Stoltenburg et al. 2005).

We sought to determine if any particular method resulted in a significant improvement for isolating high-affinity aptamers. Overall, from our data it is difficult to draw meaningful conclusions given the large bias toward certain SELEX methods. Additionally, each target type is typically associated with one or more methods. The frequency of each method by target type is shown in Fig. S11–S13. Therefore, more systematic experiments are required to fully conclude if certain methods can improve selection success and the resulting affinity of isolated aptamers. Our data and analyses are found in the Supplementary Material with a few important exceptions.

From our data, generic SELEX was by far the most commonly employed SELEX method. Approximately, 74 % of SELEX experiments were described as generic, meaning targets were of all types and a modification to the traditional approach (e.g., Flu-Mag SELEX) (Stoltenburg et al. 2005), or specialized equipment (capillary electrophoresis or microfluidics) (Berezovski et al. 2006) was not applied. Due to the bias in sample size and resultant differences in sample variance, the non-parametric Kruskal–Wallis test was used to evaluate the data. A significant difference was observed between the mean rank of Log10(K_D) between SELEX methods (Kruskal–Wallis: $\alpha = 0.05$, $p = 0.01$) which suggests that the median Log10(K_D) value between at least two methods differ (Fig. S14). More data are required to determine which methods may produce better dissociation constants.

The most widely reported partitioning method was the classic affinity chromatography method. This is where a target is chemically immobilized to a solid support matrix. 42 % of selection experiments sampled used affinity chromatography (typically agarose-based matrices) and 12 % used affinity chromatography on magnetic beads as a partitioning method. The next most frequent partitioning method used was filtration (21 %). A significant difference in the mean rank of Log10(K_D) value between partitioning methods was revealed by Kruskal–Wallis ($p \ll 0.05$) (Fig. S15). As more data become available from novel SELEX experiments, a more in-depth analysis in partitioning method-based differences at the level of target type may reveal more significant trends.

There are multiple physical and chemical methods for recovering nucleic acids. From the data sampled, heat is the most commonly used to recover nucleic acids from the bound complex (16 % of SELEX experiments). Other commonly used methods, in order of frequency, include elution with target (12 %), precipitation (11 %), chaotropic agent (11 %), and change in ionic strength (10 %). The results of a One-way ANOVA revealed a significant difference between mean Log10(K_D) for recovery methods ($\alpha = 0.05$, $p \ll 0.05$). Since the recovery method "elution with target" was used relatively frequently and also visually appeared (Fig. S16) to have a higher mean Log10(K_D), a One-way ANOVA was used to compare elution with target to other methods. There was a significant difference observed between the mean Log10(K_D) of elution with target and all other methods ($\alpha = 0.05$; $p > 0.01$ for all pairwise comparisons). This suggests that the recovery method of elution with target produces lower affinity dissociation constants. It is important to note that not all of these affinity methods are appropriate for every target type. For instance, elution with target is more typically used with small molecule targets, which tend to have lower dissociation constants in general due to the challenges associated with small molecule SELEX (McKeague and Derosa 2012). Like partitioning method, the recovery method may benefit from in-depth analysis at the level of target type.

Conclusion

Here, we investigated the effect of aptamer selection parameters and conditions on the final affinity of the resulting aptamers by comparing in vitro aptamer selection data throughout the last two decades. Specifically, we examined the properties of the aptamers (including nucleic acid type, length, and base composition), the target, the conditions (including ionic strength, buffering agent, temperature, and pH), as well as the technical methodology. From these data and our analysis, we can conclude that the nature of the target and temperature have the largest impact on the ability to obtain high-affinity aptamers. Other conditions such as Mg^{2+} concentration may be important and will be investigated further. However, currently, there is no evidence that choice of aptamer type (e.g., DNA or RNA), length, and base composition, or the selection methodology have any effect on the successful discovery of high-affinity aptamers. We anticipate though, that as more data become available, target type-specific trends will emerge. With this

information in hand, we hope that future aptamer researchers are better equipped to make informed decisions about their target of interest, template design, experimental conditions, and methodology for beginning their new selections.

Acknowledgments We thank Dr. Alex Wahba, Dr. Heather Shanks-McElroy, Ben Dorion, and Amal Awad for useful discussions in designing this research and analyzing the data. This work was supported by funds from the Natural Sciences and Engineering Research Council (NSERC) (Grant to M.C.D.) and Health Canada (Grant to R.A-R.).

Author Contributions The manuscript was written through contributions of all authors. M.M., E.M.M., M.D., and M.C.D. wrote the article.

References

Alam KK, Chang JL, Burke DH (2015) FASTAptamer: a bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. Mol Ther Nucleic acids 4:e230. doi:10.1038/mtna.2015.4

Apweiler R, Bairoch A, Wu CH (2004) Protein sequence databases. Curr Opin Chem Biol 8:76–80. doi:10.1016/j.cbpa.2003.12.004

Baird GS (2010) Where are all the aptamers? Am J Clin Pathol 134:529–531. doi:10.1309/AJCPFU4CG2WGJJKS

Barrick JE, Breaker RR (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. Genome Biol 8:R239. doi:10.1186/gb-2007-8-11-r239

Berezovski M, Musheev M, Drabovich A, Krylov SN (2006) Non-SELEX selection of aptamers. J Am Chem Soc 128:1410–1411. doi:10.1021/ja056943j

Bordeaux J et al (2010) Antibody validation. Biotechniques 48:197–209. doi:10.2144/000113382

Carothers JM, Oestreich SC, Szostak JW (2006) Aptamers selected for higher-affinity binding are not more specific for the target ligand. J Am Chem Soc 128:7929–7937. doi:10.1021/ja060952q

Carothers JM, Goler JA, Kapoor Y, Lara L, Keasling JD (2010) Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity. Nucleic Acids Res 38:2736–2747. doi:10.1093/nar/gkq082

Carter CW Jr, Yin Y (1994) Quantitative analysis in the characterization and optimization of protein crystal growth. Acta Crystallogr Sect D 50:572–590. doi:10.1107/S0907444994001228

Catherine AT, Shishido SN, Robbins-Welty GA, Diegelman-Parente A (2014) Rational design of a structure-switching DNA aptamer for potassium ions. FEBS Open Bio 4:788–795. doi:10.1016/j.fob.2014.08.008

Cho EJ, Lee JW, Ellington AD (2009) Applications of aptamers as sensors. Annu Rev Anal Chem (Palo Alto, Calif) 2:241–264. doi:10.1146/annurev.anchem.1.031207.112851

Ciesiolka J, Illangasekare M, Majerfeld I, Nickles T, Welch M, Yarus M, Zinnen S (1996) Affinity selection-amplification from randomized ribooligonucleotide pools. Methods Enzymol 267:315–335

Cruz-Toledo J et al (2012) Aptamer base: a collaborative knowledge base to describe aptamers and SELEX experiments. Database 2012:bas006. doi:10.1093/database/bas006

Deigan KE, Ferre-D'Amare AR (2011) Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs. Acc Chem Res 44:1329–1338. doi:10.1021/ar200039b

Deng Q, German I, Buchanan D, Kennedy RT (2001) Retention and separation of adenosine and analogues by affinity chromatography with an aptamer stationary phase. Anal Chem 73:5415–5421

Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346:818–822. doi:10.1038/346818a0

Ellington AD, Szostak JW (1992) Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. Nature 355:850–852. doi:10.1038/355850a0

Famulok M, Hartig JS, Mayer G (2007) Functional aptamers and aptazymes in biotechnology, diagnostics, and therapy. Chem Rev 107:3715–3743. doi:10.1021/cr0306743

Fusco D, Barnum TJ, Bruno AE, Luft JR, Snell EH, Mukherjee S, Charbonneau P (2014) Statistical analysis of crystallization database links protein physico-chemical features with crystallization mechanisms. PLoS One 9:e101123. doi:10.1371/journal.pone.0101123

Gilbert W (1986) Origin of life: the RNA world. Nature 319:618

Gold L et al (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. PloS one 5:e15004. doi:10.1371/journal.pone.0015004

Gold L, Janjic N, Jarvis T, Schneider D, Walker JJ, Wilcox SK, Zichi D (2012) Aptamers and the RNA world, past and present. Cold Spring Harb Perspect Biol. doi:10.1101/cshperspect.a003582

Hall B, Micheletti JM, Satya P, Ogle K, Pollard J, Ellington AD (2009) Design, synthesis, and amplification of DNA pools for in vitro selection. In: Ausubel FM et al (eds) Current protocols in molecular biology, Chapter 24, pp 24–22. doi:10.1002/0471142727.mb2402s88

Hianik T, Ostatna V, Sonlajtnerova M, Grman I (2007) Influence of ionic strength, pH and aptamer configuration for binding affinity to thrombin. Bioelectrochemistry 70:127–133. doi:10.1016/j.bioelechem.2006.03.012

Hoinka J, Berezhnoy A, Sauna ZE, Gilboa E, Przytycka TM (2014) AptaCluster—a method to cluster HT-SELEX aptamer pools and lessons from its application. Res Comput Mol Biol 8394:115–128. doi:10.1007/978-3-319-05269-4_9

Jagannathan V, Roulet E, Delorenzi M, Bucher P (2006) HTPSELEX–a database of high-throughput SELEX libraries for transcription factor binding sites. Nucleic Acids Res 34:D90–D94. doi:10.1093/nar/gkj049

Jayasena SD (1999) Aptamers: an emerging class of molecules that rival antibodies in diagnostics. Clin Chem 45:1628–1650

Jhaveri S, Ellington A (2002) In vitro selection of RNA aptamers to a small molecule target. In: Beaucage SL et al. (eds) Current protocols in nucleic acid chemistry, Chapter 9, pp 9–5. doi:10.1002/0471142700.nc0905s08

Lee JF, Hesselberth JR, Meyers LA, Ellington AD (2004) Aptamer database. Nucleic Acids Res 32:D95–D100. doi:10.1093/nar/gkh094

Legiewicz M, Lozupone C, Knight R, Yarus M (2005) Size, constant sequences, and optimal selection. RNA (New York, NY) 11:1701–1709. doi:10.1261/rna.2161305

Li Y, Geyer CR, Sen D (1996) Recognition of anionic porphyrins by DNA aptamers. Biochemistry 35:6911–6922. doi:10.1021/bi960038h

Lozupone C, Changayil S, Majerfeld I, Yarus M (2003) Selection of the simplest RNA that binds isoleucine. RNA (New York, NY) 9:1315–1322

Luo X et al (2010) Computational approaches toward the design of pools for the in vitro selection of complex aptamers. RNA (New York, NY) 16:2252–2262. doi:10.1261/rna.2102210

Mascini M, Palchetti I, Tombelli S (2012) Nucleic acid and peptide aptamers: fundamentals and bioanalytical aspects. Angew Chem (Int Edin English) 51:1316–1332. doi:10.1002/anie.201006630

Mattice CM, DeRosa MC (2015) Status and prospects of aptamers as drug components. BioDrugs 29:151–165. doi:10.1007/s40259-015-0126-5

McKeague M, DeRosa MC (2012) Challenges and opportunities for small molecule aptamer development. J Nucleic Acids 2012:748913. doi:10.1155/2012/748913

McKeague M, DeRosa MC (2014) Aptamers and SELEX: tools for the development of transformative molecular recognition technology. Aptamers Synth Antib 1:12–16

McKeague M, Giamberardino A, DeRosa MC (2011) Advances in aptamer-based biosensors for food safety. In: Environmental Biosensors. Vernon Somerset, InTech, SBN: 9789533074863, pp 17–18–42

McKeague M et al (2015) Comprehensive analytical comparison of strategies used for small molecule aptamer evaluation. Anal Chem. doi:10.1021/acs.analchem.5b02102

McPherson A, Cudney B (2014) Optimization of crystallization conditions for biological macromolecules. Acta Crystallogr Sect F 70:1445–1467. doi:10.1107/S2053230X14019670

Neves MA, Reinstein O, Saad M, Johnson PE (2010) Defining the secondary structural requirements of a cocaine-binding aptamer by a thermodynamic and mutation study. Biophys Chem 153:9–16. doi:10.1016/j.bpc.2010.09.009

Nimjee SM, Rusconi CP, Sullenger BA (2005) Aptamers: an emerging class of therapeutics. Annu Rev Med 56:555–583. doi:10.1146/annurev.med.56.062904.144915

Ponomarenko JV, Orlova GV, Frolov AS, Gelfand MS, Ponomarenko MP (2002) SELEX_DB: a database on in vitro selected oligomers adapted for recognizing natural sites and for analyzing both SNPs and site-directed mutagenesis data. Nucleic Acids Res 30:195–199

Price WN 2nd et al (2009) Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. Nat Biotechnol 27:51–57. doi:10.1038/nbt.1514

Robertson DL, Joyce GF (1990) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. Nature 344:467–468. doi:10.1038/344467a0

Sabeti PC, Unrau PJ, Bartel DP (1997) Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. Chem Biol 4:767–774. doi:10.1016/S1074-5521(97)90315-X

Sefah K, Shangguan D, Xiong X, O'Donoghue MB, Tan W (2010) Development of DNA aptamers using Cell-SELEX. Nat Protoc 5:1169–1185. doi:10.1038/nprot.2010.66

Serganov A, Patel DJ (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nat Rev Genet 8:776–790. doi:10.1038/nrg2172

Silverman SK (2009) Artificial functional nucleic acids: aptamers ribozymes, and deoxyribozymes identified by in vitro selection. Funct Nucleic Acids Anal Appl 1:47–108. doi:10.1007/978-0-387-73711-9_3

Stoltenburg R, Reinemann C, Strehlitz B (2005) FluMag-SELEX as an advantageous method for DNA aptamer selection. Anal Bioanal Chem 383:83–91. doi:10.1007/s00216-005-3388-9

Thodima V, Pirooznia M, Deng Y (2006) RiboaptDB: a comprehensive database of ribozymes and aptamers. BMC Bioinform 7(Suppl 2):S6. doi:10.1186/1471-2105-7-S2-S6

Tolle F, Brandle GM, Matzner D, Mayer G (2015) A versatile approach towards nucleobase-modified aptamers. Angew Chem. doi:10.1002/anie.201503652

Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505–510

Velez TE et al (2012) Systematic evaluation of the dependence of deoxyribozyme catalysis on random region length. ACS Comb Sci 14:680–687. doi:10.1021/co300111f

Waldrop MM (1989) Did life really start out in an RNA world? Science 246:1248–1249

Walsh R, DeRosa MC (2009) Retention of function in the DNA homolog of the RNA dopamine aptamer. Biochem Biophys Res Commun 388:732–735. doi:10.1016/j.bbrc.2009.08.084

Zhu G, Ye M, Donovan MJ, Song E, Zhao Z, Tan W (2012) Nucleic acid aptamers: an emerging frontier in cancer therapy. Chem Commun 48:10472–10480. doi:10.1039/c2cc35042d