Evidence of In Silico Unexplained Aptamer Sequence Alterations Demands Collaborative Evidence-Based Aptamer Publication Guidelines

Alexandra A. Miller, Abhijit Rao, Sujana Nelakanti, Christopher Kujalowicz, Ted Shi, Ted Rodriguez, Andrew D. Ellington, Gwendolyn M. Stovall

LOYOLA UNIVERSITY CHICAGO HEALTH SCIENCES DIVISION

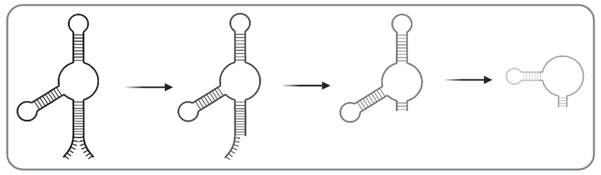
FRESHMAN RESEARCH INITIATIVE
COLLEGE OF NATURAL SCIENCES

Introduction

Aptamers are short, single-stranded nucleotide sequences that can bind to a target molecule with high specificity, allowing for a robust range of industrial, diagnostic, clinical, and therapeutic applications. Aptamers have been the subject of more than 144,000 papers to date. There has been a growing concern that discrepancies in the reporting of aptamer research limit the reliability of these reagents for research and other applications.

We first noted the misinterpretation of an 80-mer RNA anti-lysozyme aptamer. Subsequent works not only altered the aptamer sequence, but also truncated the aptamer. We hypothesized that the alteration of the anti-lysozyme aptamer may not have been an isolated event.

These observations noting inconsistencies in the use of the RNA anti-lysozyme aptamer served as an impetus for our systematic review of the reporting of aptamer sequences in the literature. Our purpose was to determine the depth and breadth of aptamer sequence infidelity throughout the literature.



in silico unexplained aptamer sequence alterations

Methods

An anti-lysozyme aptamer in 2001 had been reported to ‘mutate’ in the literature over time, and we therefore used this aptamer as a starting point for establishing a broader methodology.

We first identified an originating aptamer and then examined the literature citing the sequence of this originating aptamer. In cases where multiple clones selected in a paper were used in the literature, all clones were examined as “root” aptamers. Unexplained sequence alterations were again categorized, and a phylogeny was constructed to document the evolution of these alterations.

Following our analysis of the anti-lysozyme aptamer, a review of aptamers against the ten most used targets was performed. This review systematically sampled the aptamer literature, spanning years, researchers, labs, and locations.

In all, 780 publications from 23 originating/root aptamers were reviewed using this standardized sampling methodology.

Of those reports that varied from the original RNA aptamer, the papers were reviewed to ascertain if the reported alterations were adequately described, omitted, or unexplained. Sequence alterations found include deletions, insertions, and/or substitutions. Samples ranged from 9-171 publications, although we aimed to acquire at least 50 papers for each phylogeny.

Results

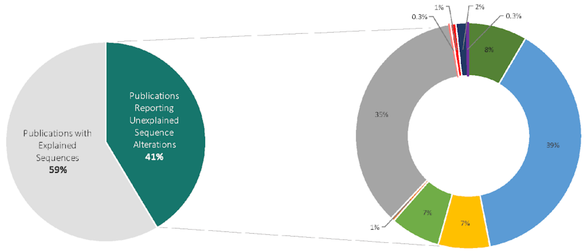
Our detailed examination of literature citing the RNA anti-lysozyme aptamer revealed that 93% of the 61 publications reviewed reported unexplained altered sequences with 96% of those using DNA variants.

We expanded our search to analyze each of the ten most cited aptamers, where we discovered that 41% of the 780 aptamer publications we reviewed reported unexplained sequence alterations.

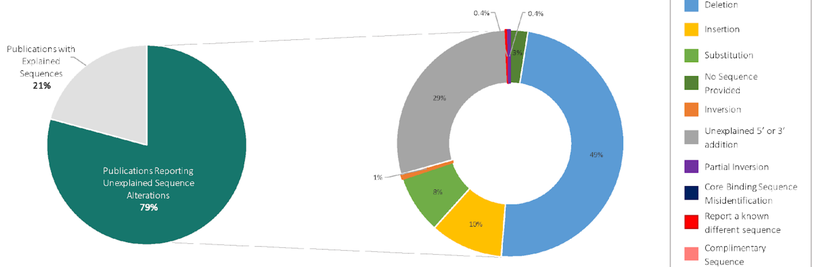
Overall, only 59% of the 780 publications our group reviewed correctly reported the aptamer sequence(s) and/or explained sequence alterations, while 41% contained one or more *in silico* sequences that were categorized as unexplained. We identified 120 novel sequences that according to our criteria were not adequately explained. 39% of these 120 sequences identified contained deletions, 35% contained 5’ or 3’ unexplained additions, 8% did not provide the sequence at all, 7% contained insertions, 7% contained substitutions, and less than 2% contained core binding sequence misidentifications, inversions, the complementary sequence, or an entirely different sequence.

The figure below illustrates the distribution of sequence alterations throughout the papers our group reviewed, including further stratification into high and low apparent error phylogenies.

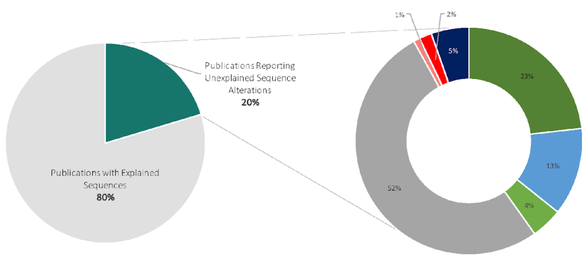
A. Distribution of Unexplained Sequence Alterations



B. Distribution of Unexplained Sequence Alterations in High Apparent Error Phylogenies (>34%)



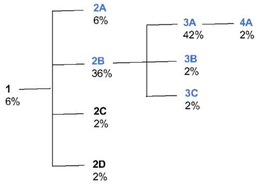
C. Distribution of Unexplained Sequence Alterations in Low Apparent Error Phylogenies (≤34%)



Discussion

The figure below illustrates the results of our anti-lysozyme aptamer analysis, the methodology of which was used throughout our investigation. Clones were analyzed for fidelity to the original sequence, alterations were described and documented, and a phylogeny detailing the evolution of sequence alterations over time was constructed.

A. Phylogeny Organization of Unexplained Anti-Lysozyme Aptamer Sequence Alterations



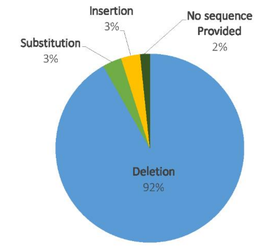
B. Suspected Source of Node 3A: Figure 3 Reporting of Clone 1[7]

Clone 1.	ATCAGGGCTAAAGAGTGCAGAGTTACTTAG	20/33 clones = 61% of sequences
Clone 2.	GGTGATCATGGCAGTGTACGCGGGCGGACA	4/33 clones = 12% of sequences
Clone 3.	GGTTGTGAAGATTGGGAGCGTCGTGGCTAC	3/33 clones = 9% of sequences
Clone 4.	GTAAATCGTCGACAGGAATTGGCGGGCCGG	3/33 clones = 9% of sequences
Clone 5.	GAATTGCGACAGTCGGGACATGTCGCGAGG	2/33 clones = 6% of sequences
Clone 6.	GGAATGAGTGCCCTGCAAGCGAGGGCTAGC	1/33 clones = 3% of sequences

C. Suspected Source of Node 2B: Table 1 Reporting of Clone 1[12]

	sequences
clone 1	5’GGGAATGGATCCACATCTACGAATTCATCAGGGCTAAAGAG TGCAGAGTTACTTAGTTCACTGCAGACTTGACGAAGCTT
clone 2	5’GGGAATGGATCCACATCTACGAATTCGGTGATCATGGCAGT GTACGCGGCGGACATTCACTGCAGACTTGACGAAGCTT
clone 3	5’GGGAATGGATCCACATCTACGAATTCGGTTGTGAAGATTGG GAGCGTCGTGGCTACTTCACTGCAGACTTGACGAAGCTT
clone 4	5’GGGAATGGATCCACATCTACGAATTCGTAAATCGTCGACAG GAATTGGCGGGCCGGTTCACTGCAGACTTGACGAAGCTT
clone 5	5’GGGAATGGATCCACATCTACGAATTCGAATTGCGACAGTCG GGACATGTCGCGAGGTTCACTGCAGACTTGACGAAGCTT
clone 6	5’GGGAATGGATCCACATCTACGAATTCGGAATGAGTGCCCTG CAAGCGAGGGCTAGCTTCACTGCAGACTTGACGAAGCTT
clone E	5’GGGAAUGGAUCCACAUCUACGAAUUCGGAACCCGCUAGGA GUGAAACGGGGCGUGCUUCACUGCAGACUUGACGAAGCUU
clone J	5’GGGAAUGGAUCCACAUCUACGAAUUCGUUGGAGAAUAUUA GCGCUACAGAACAGGUUCACUGCAGACUUGACGAAGCUU
clone M	5’GGGAAUGGAUCCACAUCUACGAAUUCGGAAAUUGGGGUUA UUGUUGCAACGGACGGUUCACUGCAGACUUGACGAAGCUU

D. Distribution of Unexplained Anti-Lysozyme Sequence Alterations



Conclusion

Our literature search bolsters the argument that the field of aptamer research is experiencing considerable inconsistencies in aptamer sequence reporting. Further, these widespread inconsistencies warrant the application of collaborative, evidence-based aptamer publication guidelines to improve reproducibility and consistency within the literature.

In the future, collaborative evidence-based aptamer validation guidelines should be encouraged by journals, including a checklist for the peer review process. Ultimately, we believe that the standardization of aptamer publication guidelines and increased availability of raw data will lead to a more open, nuanced discussion of the data presented, and greater success in the translation of aptamer research to the clinic and industry.

Our findings can be used as a starting point for building better practices in author submissions and publication standards, the rigor, and reproducibility of aptamer research.

At a minimum, we suggest that three categories of information should be required in aptamer publications: 1. complete sequence and secondary structure information (i.e., mFOLD), including consistent sequence identifiers; 2. detailed descriptions of binding and experimental conditions; and 3. extensive use of negative and positive controls, for both aptamers and targets.