

# Systematic Review of Aptamer Sequence Reporting in the Literature Reveals Widespread Unexplained Sequence Alterations

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



Cite This: *Anal. Chem.* 2022, 94, 7731–7737



Read Online

ACCESS |



Metrics & More

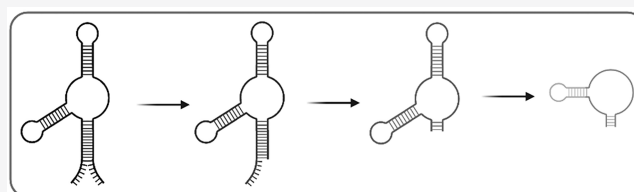


Article Recommendations



Supporting Information

**ABSTRACT:** Aptamers have been the subject of more than 144 000 papers to date. However, 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. These observations noting inconsistencies in the use of our RNA antilysozyme aptamer served as an impetus for our systematic review of the reporting of aptamer sequences in the literature. Our detailed examination of the literature citing the RNA antilysozyme aptamer revealed that 93% of the 61 publications reviewed reported unexplained altered sequences with 96% of those using DNA variants. The 10 most cited aptamers were examined using a standardized methodology in order to categorize the extent to which the sequences themselves and altered sequences were adequately described in the literature. Our review of 780 aptamer publications spanned decades, multiple journals, and research groups and revealed that 41% of the papers reported unexplained sequence alterations or omitted sequences. We identified 10 common categories of sequence alterations including deletions, substitutions, and additions, among others. Overall, our findings can be used as a starting point for building better practices in author submissions and publication standards, elevating the rigor and reproducibility of aptamer research.



in silico unexplained aptamer sequence alterations

The irreproducibility of research has become increasingly evident in several fields in recent years<sup>1</sup> and poses a serious threat to the efficacy and validity of the research and the general public's perception of science. The field of aptamer research may be particularly susceptible to this reproducibility crisis.<sup>2–5</sup> In part, the ability to develop affinity reagents quickly and simply based on molecular biology manipulations alone provides a ready entry point for a variety of researchers with varying backgrounds. The development of aptamers for use stands in stark contrast to the development of monoclonal antibodies via hybridoma and other technologies, where there is a much larger and longer trail of publications and standards, both academic and commercial, and where there is much larger use, allowing greater cross-validation. The reliability of aptamers has at times been drawn into question.<sup>3,6</sup>

As an early example of the issues relating to the reproducibility and reliability of aptamer research, we first noted the misinterpretation of an 80-mer RNA antilysozyme aptamer (clone 1).<sup>2,7</sup> As we now report in greater detail, subsequent works not only altered the aptamer sequence (using DNA as opposed to the original RNA aptamer) but also truncated the aptamer.<sup>4,8</sup>

We hypothesized that the sequence alteration of the antilysozyme aptamer may not have been an isolated event. To better elucidate the extent of *in silico* aptamer sequence alterations, we sampled a large swath of aptamer literature to gauge the fidelity of aptamer sequence reporting. We examined

aptamers against the most frequently used aptamer targets: thrombin, adenosine triphosphate (ATP), vascular endothelial growth factor (VEGF), platelet derived growth factor (PDGF) BB, cocaine, theophylline, lysozyme, nucleolin, immunoglobulin E (IgE), and ochratoxin A (OTA; top 10 targets listed by application-based publication frequency).<sup>9</sup> We identified original aptamer sequences and followed their subsequent descriptions in cited literature, characterizing apparent sequence alterations as adequately explained, omitted, or unexplained. The unexplained aptamer sequence findings were then organized into phylogenetic trees that aided in our identification of common sources and types of apparent mutation.

In our review of 780 publications, we provide evidence of widespread unexplained aptamer sequence alteration over time. Aptamers appear to “mutate” in the literature.

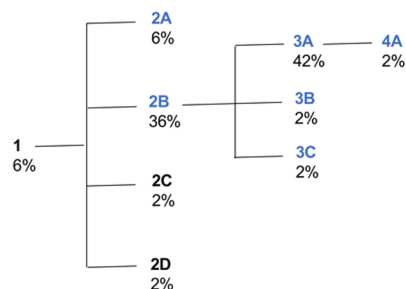
**Received:** October 10, 2021

**Accepted:** March 24, 2022

**Published:** April 14, 2022



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

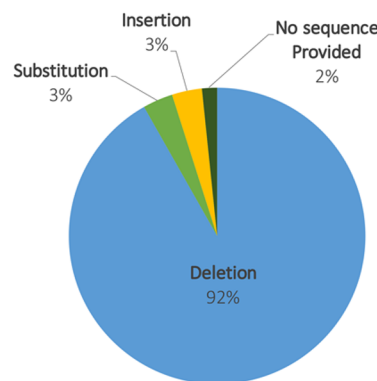
B. Suspected Source of Node 3A: Figure 3 Reporting of Clone 1<sup>7</sup>

<b>Clone 1.</b>	ATCAGGGCTAAAGAGTGCAGAGTTACTTAG 20/33 clones = 61% of sequences
<b>Clone 2.</b>	GGTGATCATGGCAGTGTACGCGGGCGGACA 4/33 clones = 12% of sequences
<b>Clone 3.</b>	GGTTGTGAAGATTGGGAGCGTCGTGGCTAC 3/33 clones = 9% of sequences
<b>Clone 4.</b>	GTAATCGTCGACAGGAATTGGCGGGCCGG 3/33 clones = 9% of sequences
<b>Clone 5.</b>	GAATTGCGACAGTCGGGACATGTCGCGAGG 2/33 clones = 6% of sequences
<b>Clone 6.</b>	GGAATGAGTGCCTGCAAGCGAGGGCTAGC 1/33 clones = 3% of sequences

C. Suspected Source of Node 2B: Table 1 Reporting of Clone 1<sup>12</sup>

	sequences
clone 1	5'GGGAATGGATCCACATCTACGAATTCATCAGGGCTAAAGAG TGCAGATTACTTAGTTCTACGCAGACTTGACGAAGCTT
clone 2	5'GGGAATGGATCCACATCTACGAATTCGGTGTATCGGCAGT GTACGCGGGCGGACATTCAGTGCAGACTTGACGAAGCTT
clone 3	5'GGGAATGGATCCACATCTACGAATTCGGTGTGAAGATTGG GAGCGTCGTGGCTACTTCTACGCAGACTTGACGAAGCTT
clone 4	5'GGGAATGGATCCACATCTACGAATTCGTAATCGTCGACAG GAATTGGCGGGCCGGTTCAGTGCAGACTTGACGAAGCTT
clone 5	5'GGGAATGGATCCACATCTACGAATTCGGAATGCGACAGTCG GGACATGTGCGGAGGTTCTACGCAGACTTGACGAAGCTT
clone 6	5'GGGAATGGATCCACATCTACGAATTCGGAATGAGTGCCTG CAAGCGAGGGCTAGTTCAGTGCAGACTTGACGAAGCTT
clone E	5'GGGAUUGGAUCCACAUUCGAAUUCGGAACCCGCUAGGA GUGAAACGGGGCGUGUUCACUGCAGACUUGACGAAGCUU
clone J	5'GGGAUUGGAUCCACAUUCGAAUUCGGAUUGGAAUUA GCGCUACAGAACAGGUUCACUGCAGACUUGACGAAGCUU
clone M	5'GGGAUUGGAUCCACAUUCGAAUUCGGAUUGGGUUA UUGUUGCAACGGACGGUUCACUGCAGACUUGACGAAGCUU

## D. Distribution of Unexplained Anti-Lysozyme Sequence Alterations



**Figure 1.** Phylogeny depicting unexplained sequence alterations introduced to the clone 1 anti-lysosome RNA aptamer<sup>7</sup> collected 01/01/2020. (A) The nodes in the artificial phylogenetic trees represent groupings of aptamers according to unexplained sequence alterations. (B) Reprinted with permission ref 7. Copyright 2001 Elsevier. (C) Reprinted in part with permission from ref 12. Copyright 2004 American Chemical Society. (D) Unexplained sequence alterations were classified into categories.

## INFORMATICS METHODOLOGY

**Examination of the Reported Sequences for the RNA Anti-Lysozyme Aptamer, Clone 1.**<sup>7</sup> An antilysozyme aptamer selected in 2001 had previously been reported to “mutate” in the literature over time, and we therefore used this aptamer as a starting point for establishing a broader methodology.<sup>2,4</sup> As part of this methodology, we identified publications by using Google Scholar and identifying papers that cited the original work using the “cited by” option for this work. From the top literature results, those that were written in English, primary literature (i.e., not reviews, etc.), and used the aptamer *in vitro* or *in vivo* were analyzed for reported sequence information. The aptamer sequence(s) was (were) extracted with notes on any alterations to the sequence, using direct quotes from the text.

Analysis of the collected sequences then included the following:

1. Comparison of the reported sequences to the original aptamer sequence (Clone 1)<sup>7</sup> and identification of alterations.

2. If alterations were present, further examination of the cited publication led to an assessment of whether the variants were adequately explained or unexplained.

3. In those cases where there were unexplained sequence alterations, they were classified into 10 categories: (i) deletion, (ii) insertion, (iii) substitution, (iv) complete inversion, (v) 5′/3′ addition, (vi) partial inversion, (vii) complementary sequence, (viii) core binding sequence misidentification (i.e.,

in a figure or table), (ix) sequence omitted, and (x) unknown (i.e., reported an entirely different sequence).

Using phylogenetic trees as a model, we organized the aptamers reported in the literature similarly, using sequence homology to visualize sequence fidelity in our sample. We loosely dubbed this organization an aptamer “phylogenetic” analysis. The nodes in the artificial phylogenetic trees represent groupings of aptamers according to unexplained sequence alterations. Further unexplained derivations were depicted as branching from a previous sequence alteration (i.e., Figure 1, node 4A was derived from Figure 1, node 3A). Publications that correctly reported the original aptamer sequence, including those with fully explained or described additions (e.g., 5′-terminal polyT linker or primary amine for conjugation), were given the same identifying number as the original aptamer.

**Examination of Aptamer Sequences Across the Literature.** To provide broader insights into the fidelity of aptamer sequence reporting, we reviewed aptamers against the 10 most-used targets<sup>9</sup> (Table S1) and again determined a primary (original) aptamer sequence and identified derivatives using the Google Scholar “cited by” option (i.e., works citing the original). From the results generated, publications were further sieved by identifying only those that reported aptamers against the target using the “search within citing articles” function and the search terms “[target]” “aptamer” (Table S1). Without this second search, extraneous papers—for example, those citing the anti-VEGF aptamer in the Introduction—

Table 1. Unexplained Sequence Alterations Introduced to the Clone 1 Anti-Lysosome RNA Aptamer<sup>7a</sup>

node	#	aptamer sequence reported
1	4	GGGAAUGGAUCCACAUCUACGAAUUAUCAGGGCUAAAAGAGUGCAGAGUUACUUAGUUCACUGCAGACUUGACGAAGCUU
2A	4	biotin-24polyT-GGGAATGGATCCACATCTACGAATTCATCAGGGCTAAAGAGTGCAGAGTTACTTACTTGCAGACTTCACGAAGCTT
2B	20	<del>GGGAATGGATCCACATCTACGAATTCATCAGGGCTAAAGAGTGCAGAGTTACTTACTTGCAGACTTCACGAAGCTT</del>
2C	1	GGGAAUGGAUCCACAUCUACGAAUUAUCAGGGCUAAAAGAGTGCAGAGUUACUUAGUUCACUGCAGACUUGACGAAGCUU
2D	1	GGGAAUGGAUCCACAUCUACGAAUUAUCAGGGCUAAAAGAGUGCAGAGUUACUUAGUUCACUGCAGACUUGACGAAGCUU
3A	25	<del>GGGAATGGATCCACATCTACGAATTCATCAGGGCTAAAGAGTGCAGAGTTACTTACTTGCAGACTTCACGAAGCTT</del>
3B	1	HS-(CH <sub>2</sub> ) <sub>6</sub> - GGGAATGGATCCACATCTACGAATTCATCAGGGCTAAAGAGTGCAGAGTTACTTACTTGCAGACTTCACGAAGCTT
3C	1	<del>GGGAATGGATCCACAGTGTATCTACGAATTCATCAGGGCTAAAGAGTGAAGAGTTACTTACTTGCAGACTTCACGAAGCTT</del>
4A	1	<del>GGGAATGGATCCACATCTACGAATTC</del> dithiolTTTTTTATCAGGGCTAAAGAGTGCAGAGTTACTTACTTGCAGACTTCACGAAGCTT
0	2	no sequence in text or supplement

<sup>a</sup>Unexplained insertions are bold. Unexplained substitutions are bold and underlined. Unexplained deletions are struck out, and justified or explained alterations are in light grey. The number (#) column indicates the number of publications found reporting each sequence in our analysis.

would have been included across multiple reviews or other experimental papers. We then categorized the unexplained sequence alterations according to apparent type and organized the reported sequence derivatives into “phylogenies” according to sequence homology.

The literature search for each aptamer excluded publications that did not use the aptamer sequence experimentally or that were not in the English language; however, no exclusions were made based on the type of journal or date of publication. Publications were searched from some of the original selection experiments in 1990<sup>10,11</sup> through March 2020.

The sample size for each aptamer examined varied between 9 and 171 publications, although we aimed to acquire at least 50 papers for each phylogeny.

## RESULTS AND DISCUSSION

### A Study of the RNA Anti-Lysozyme Aptamer, Clone 1.<sup>7</sup>

Lysozyme is one of the most utilized targets in the field of aptamer research and is among the top 10 in terms of citations.<sup>9</sup> This popularity was driven in large measure by the publication of an RNA antilysozyme aptamer by Cox et al.<sup>7</sup> We previously noted<sup>2</sup> that while the original selection had been for an RNA aptamer, subsequent uses of this aptamer had sometimes involved the synthesis of a DNA version that had not been originally verified as binding to lysozyme; in other words, the aptamer had been mutated *in silico*. This chemical mutation was perhaps understandable, given that the figure in the original paper showed the DNA version of the RNA aptamers without the primer regions (although the text clearly describes the generation of RNA transcripts and Figures 2 and 4 note the use of RNA). This misreading and misattribution of the original work led to the *in silico* alterations propagated by subsequent groups.

To examine how *in silico* alterations can be propagated, we reviewed 61 primary research papers that cited the antilysozyme aptamer, clone 1.<sup>7</sup> 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.

In an effort to illustrate and organize the aptamer sequence data as it “evolved” in the literature, an aptamer phylogeny was created. The phylogeny includes the original/parental 80-mer antilysozyme RNA aptamer (clone 1),<sup>7</sup> also called the “root” aptamer, and nodes are organized by sequence homology (i.e., those that more greatly differ from the root aptamer are further

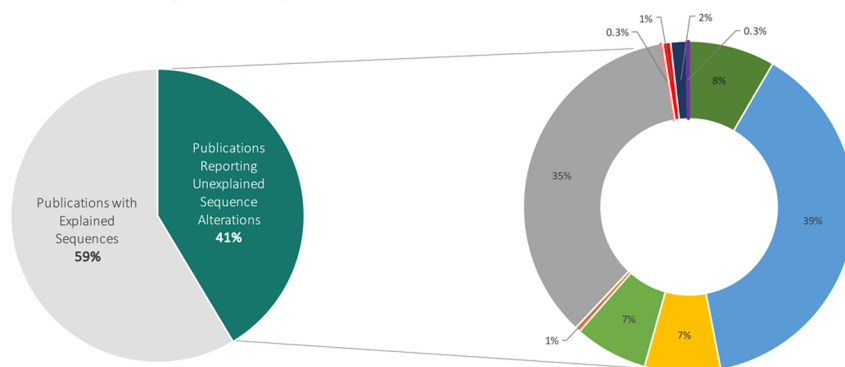
away and have a larger node number). As shown in Figure 1A, the phylogeny includes four branch points and nine nodes in all (including the root aptamer) with DNA variants indicated as blue nodes. While we made our best attempts to ascertain what was “reported” versus what was actually “used,” in most cases, we could not distinguish one from the other and thus settled on the liberal use of what was “reported.” Further, in an attempt to ascertain the source of the unexplained sequence alterations, we examined altered sequences and the publications they cited and report our hypotheses here. The sequence alignment of unexplained alterations of the antilysozyme aptamer is shown in Table 1.

After the root antilysozyme RNA aptamer node (1) in Figure 1, the first branch point leads to four nodes. The first node, 2A, is an unexplained truncated DNA sequence with a 39 nt 3′ deletion, as well as terminal additions (5′ biotinylation and a 24 nt oligo(T) linker). This aptamer was primarily used by a single group.<sup>13–16</sup> We hypothesize that the alteration was due to a misinterpretation of Kirby et al. (Table 1),<sup>12</sup> shown in Figure 1C, which reported the DNA version of the six original antilysozyme aptamers<sup>7</sup> and three novel RNA aptamers with a line break in the sequences. Node 2A includes the first line of this sequence.

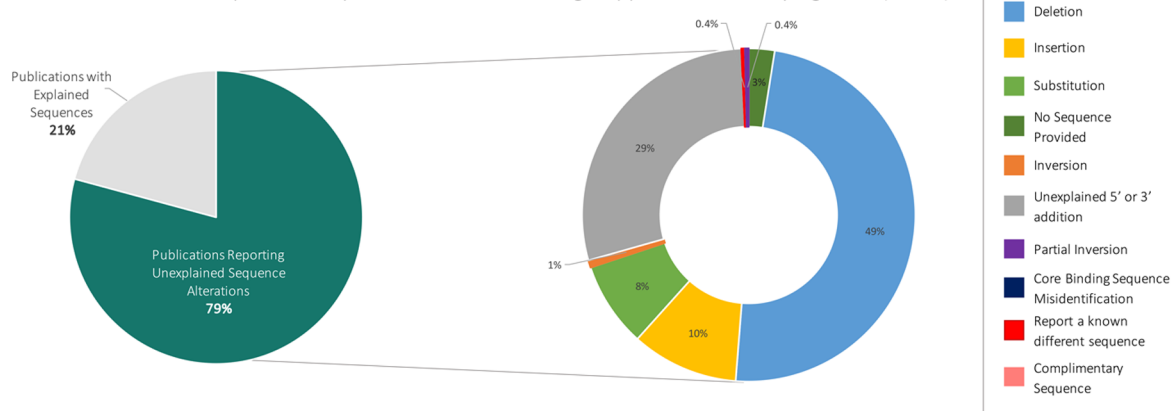
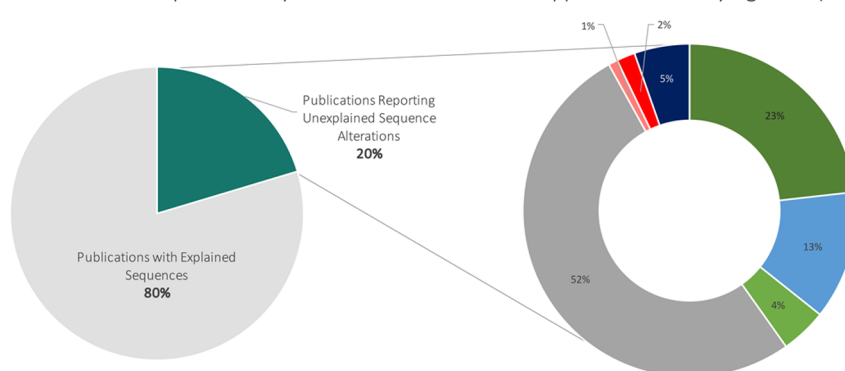
The 2B node (Figure 1) is an unexplained DNA sequence with a 24 nt deletion from the 3′ end and a 14 nt deletion from the 5′ end and was found in 36% of papers reviewed. The 2C node is the RNA aptamer sequence, but with a single U to T substitution reported in the paper, an *in silico* error but likely not in the molecule used, as this is not described in the Methods section.<sup>17</sup> Finally, the 2D node is a full-length RNA sequence but with a deletion of four internal nucleotides,<sup>18</sup> likely again an error in the description but not in the molecule itself, as the manuscript notes the acquisition of the original aptamer material from the lab where the aptamer originated.

All further unexplained sequence alterations branch from node 2B, the citation of an all-DNA variant (Figure 1A). Organized by sequence homology, the 5′ and 3′ cropped variant at the 2B node leads to a branch point with three nodes: 3A, 3B, and 3C. Node 3A, which makes up 42% of the altered sequences in this phylogeny, is a DNA variant with primer-binding regions deleted (i.e., 5′ 26 nt deletion and 3′ 24 nt deletion). This sequence is equivalent to what was reported in Figure 3 of the 2001 paper,<sup>7</sup> although this sequence was not in fact used for experiments. We suspect that the node 3A sequence alteration arose due to a misunderstanding of Figure 3 in the original paper, which reported a DNA version of the 30 nt random sequence region of the

## A. Distribution of Unexplained Sequence Alterations



## B. Distribution of Unexplained Sequence Alterations in High Apparent Error Phylogenies (&gt;34%)

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

**Figure 2.** Distribution of field-wide unexplained aptamer sequence alterations. (A) The breakdown of all publications examined ( $n = 780$  publications, 23 phylogenies grouped by 11 root aptamer publications). (B) The distributions of phylogenies that contained greater than the median of 34% unexplained internal sequence alteration (five phylogenies: Cox lysozyme, lysozyme (DNA aptamer), ATP, PDGF BB, cocaine). (C) The distribution of phylogenies that contained less than or equal to the median of 34% internal mutations (six phylogenies: nucleolin, IgE, theophylline, VEGF, ochratoxin A, thrombin). The percentage of unexplained sequence alterations within each phylogeny can be found in Table S2.

selected RNA aptamer, without also including accompanying primer-binding regions.<sup>7</sup> Presumably, groups that used the DNA version of the random region obtained it by using this figure.

Variant 3B differs from 3A in that it has an unexplained 3'-terminal addition of GGGCGC, the partial (13 nt) return of the 5' primer region, and an adequately explained 5'-thiol modification ( $S'$ -HS-( $CH_2$ )<sub>6</sub>) for covalent conjugation to gold nanoparticles.<sup>19</sup> Variant 3C contains a number of otherwise unexplained alterations: a C to A internal substitution, a 6 nt 5'-terminal addition, a 3 nt 3'-terminal deletion, and a 5'-terminal 3 nt deletion.<sup>20</sup>

Finally, the 4A sequence (Figure 1) again builds off of the deletion of 5' and 3' primer regions described in node 3A and adds a well-described 6 nt polyT linker and thiol modification at the 5' end for conjugation to silica particles but also includes an unexplained additional 13 nt 3' deletion.<sup>21</sup>

In all of the reviewed publications, there were only four papers of the 61 reviewed that reported the correct antilysozyme aptamer, clone 1, sequence and/or adequately explained sequence alterations. Of those that contained unexplained sequence alterations, 96% reported a DNA variant, 92% contained a deletion, 3% contained a substitution,

3% contained an insertion, and 2% failed to include sequence information in the manuscript or supplement.

The question thus becomes not only why so many researchers misinterpreted the primary source of information but also how the derivatives themselves work as reported. The answer is two-fold: first, hen egg-white lysozyme was originally chosen as a target for selection in part because it is extremely basic ( $pI = 11.35$ ) and thus was very likely to yield high-affinity aptamers. But second, precisely because polyanions will bind oligocations, the DNA aptamer may still bind lysozyme. In an illuminating study, Potty et al.<sup>4</sup> examined the binding properties of both the RNA aptamer and DNA of antilysozyme aptamer clone 3 (an entirely different aptamer sequence than the clone 1 sequence that is generally used). These authors found the DNA variant bound nonspecifically to lysozyme through electrostatic interactions.<sup>4</sup>

This further raises the question of what constitutes specific binding. In general, aptamer studies should contain non-cognate binding experiments, in which variants of the aptamers (mutant or scrambled) are shown not to bind to the selection target, and the selected aptamer is shown not to bind to targets other than the selection target. Such studies are rarely carried out.<sup>9</sup>

**Description of Aptamer Sequence Reporting Across the Literature.** Given these findings, we broadly sampled the aptamer literature to more fully describe sequence reporting fidelity in the field. On the basis of the extensive set of unexplained aptamer sequence alterations that were found in the literature citing the antilysozyme aptamer,<sup>7</sup> we hypothesized that the alteration of original aptamer sequences and/or unexplained sequence alterations could be a pervasive problem. While a full review of the 144 000+ “aptamer” papers on Google Scholar was not feasible, a review of aptamers against the 10 most used targets<sup>9</sup> was carried out. This Review systematically sampled a broad swath of the aptamer literature, spanning years, researchers, laboratories, and locations.

Similar to the analysis of the Cox antilysozyme aptamer literature, 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 (23 in all). Unexplained sequence alterations were again categorized, and phylogenies constructed. In all, 780 publications from 23 originating/root aptamers were reviewed using this standardized sampling methodology (Table S2, Figures S1–10). The 23 phylogenies were also grouped based on an originating publication (e.g., three different anti-PDGF aptamer clones from the originating publication are grouped in Figure S4).

Overall, only 59% of the 780 publications 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 (Table S3). Some (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 (i.e., boxing the incorrect aptamer sequence, etc.), inversions, the complementary sequence, or an entirely different sequence (Figure 2A).

To further categorize the types of apparent *in silico* mutations that are perpetuated in the literature, we used a “median split” to divide the phylogenies into two groups: “low”/few apparent mutations and “high”/more apparent mutations.<sup>22,23</sup> The median percentage of unexplained sequence alterations within each phylogeny was 34%, and this value was used for the division into groups (Figure 2B and C, Tables S2, S4). There was a larger contribution of deletions (49%), insertions (10%), and substitutions (8%) in the high level of unexplained sequence alteration group relative to the low level of the unexplained sequence alteration group, which had 13% deletions, 4% substitutions, and no insertions (Figure 2B and C). In contrast, the low-level unexplained sequence alteration group contained more unexplained 5′/3′ additions (52%) compared to the high sequence alteration group (29%). This observation generally suggests that deletions and insertions are further propagated in subsequent literature, while 5′/3′ additions that may have a unique, researcher-specific purpose do not. It has previously been noted that many aptamers appear to not utilize flanking sequences during their folding,<sup>24</sup> and many 5′/3′ extensions may therefore not have an impact on function.

Since this Perspective describes *in silico* mutations of aptamers, it is primarily meant as a cautionary tale for authors, rather than as an extensive biophysical or computational characterization of the impact of those *in silico* mutations. The lack of a comprehensive, updated aptamer database is clearly an impediment to even something as straightforward and potentially useful as analyzing aptamer  $K_d$  values. Gathering such data across the literature will require enormous time and analytical commitments, given the extremely diverse way in which aptamer data are already reported (for example, nominal  $K_d$  values are typically gathered by a variety of methods and calculated in a variety of ways). However, creating a unified database is something the field as a whole or companies engaged in aptamer research might wish to promote. Overall, sequence data provide the best and currently only reliable means to make comparisons across 30 years of literature.

## CONCLUSIONS

Our review of 780 publications reveals widespread sequence alterations that are poorly explained or where the aptamer sequence(s) were omitted (41%). This observed systemic mutation of aptamers via the literature buttresses the conclusions of Yan and Levy<sup>3</sup> that aptamers can potentially prove unreliable as reagents due to author error. Other reports have raised issues in aptamer research that may similarly impact the propagation of literature mutations relative to originating sequences—the relative ease of creating aptamers based on literature reports alone, the rapid expansion of aptamer research, the rapid increase in application-based publications, and a lack of adherence to publication guidelines.

Echoing previous work,<sup>2,3,9</sup> we assert that there is a need both for the collaborative construction of evidence-based publication guidelines and for standardized documentation of aptamer sequence use in the literature (e.g., similar to CiteAb.com). Standardized sequence reporting could be modeled after Nature’s “Life Science Reporting Summary” and Cell’s “STAR Methods,” which create standardized templates for detailed descriptions of experimental design, statistical tests, and materials and reagents. At a minimum, we suggest that three categories of information should be required in aptamer publications: (1) complete sequence, 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. In addition, while aptamer secondary structures are frequently provided based on generally available folding algorithms such as mFOLD,<sup>25</sup> authors should generally state that any secondary structures that have been identified computationally are not necessarily accurate in the absence of further experimental characterizations. As Yan and Levy<sup>3</sup> suggest, we also support validation practices such as using conditions similar to those found in downstream applications, limiting incubation times and using blocking reagents to minimize nonspecific binding to targets, and performing multiple validations, ideally by different researchers.

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.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.1c04407>.

(Table S1) Top 10 targets used and the search terms used during the Google Scholar literature search; (Table S2) percentage of publications with unexplained sequence alterations within each phylogeny; (Table S3) number of unique sequences and number of sequences with unexplained alterations; (Figures S1–S10) extent of publication-induced mutation in the top 10 aptamer targets using the method described here; (Tables S4 and S5) the composite data (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

**Gwendolyn M. Stovall** – *Texas Institute for Discovery Education in Science Freshman Research Initiative and Texas Institute for Discovery Education in Science High School Research Initiative, The University of Texas at Austin, Austin, Texas 78712, United States*; [orcid.org/0000-0002-7525-9733](https://orcid.org/0000-0002-7525-9733); Email: [gwenstovall@utexas.edu](mailto:gwenstovall@utexas.edu)

### Authors

**Alexandra A. Miller** – *Texas Institute for Discovery Education in Science Freshman Research Initiative, The University of Texas at Austin, Austin, Texas 78712, United States*

**Abhijit S. Rao** – *Texas Institute for Discovery Education in Science Freshman Research Initiative, The University of Texas at Austin, Austin, Texas 78712, United States*

**Sujana R. Nelakanti** – *Texas Institute for Discovery Education in Science Freshman Research Initiative, The University of Texas at Austin, Austin, Texas 78712, United States*

**Christopher Kujalowicz** – *Texas Institute for Discovery Education in Science Freshman Research Initiative, The University of Texas at Austin, Austin, Texas 78712, United States*

**Ted Shi** – *Texas Institute for Discovery Education in Science Freshman Research Initiative, The University of Texas at Austin, Austin, Texas 78712, United States*

**Ted Rodriguez** – *Texas Institute for Discovery Education in Science Freshman Research Initiative, The University of Texas at Austin, Austin, Texas 78712, United States*

**Andrew D. Ellington** – *Institute for Molecular Biosciences and Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, Texas 78712, United States*;

[orcid.org/0000-0001-6246-5338](https://orcid.org/0000-0001-6246-5338)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.analchem.1c04407>

### Author Contributions

A.A.M., S.R.N., A.S.R., C.K., T.R., and T.S. performed initial data collection. A.A.M., S.R.N., and A.S.R. collected all data and analyzed the data. A.A.M., G.M.S., and A.D.E. reviewed and interpreted the results. The methods were drafted by S.R.N. The final manuscript was written and edited by A.A.M., G.M.S., and A.D.E. All authors have approved the final version of the manuscript.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research was funded by The University of Texas Freshman Research Initiative, which was supported by Howard Hughes Medical Institute (#52008124) and the College of Natural Sciences. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of The University of Texas at Austin or the Howard Hughes Medical Institute. Apart from Gwendolyn Stovall and Andrew Ellington, the authors began this work as undergraduate researchers in the UT Freshman Research Initiative and would like to thank the FRI's Aptamer Stream and the 2018 TIDES Science Sprints at the University of Texas at Austin for supporting this project. The authors would also like to thank Jolyn Frnka, Haley Wolf, Shreya Thiagarajan, Eric Lumanog, Christine Mai, Hung Le, Kathy Tran, and Nidhi Pawate for their assistance in the cursory data collection and development of the method.

## ■ REFERENCES

- (1) Miyakawa, T. *Molecular Brain* **2020**, *13* (1), 1–6.
- (2) Cho, E. J.; Lee, J.-W.; Ellington, A. D. *Annual Review of Analytical Chemistry* **2009**, *2* (1), 241–264.
- (3) Yan, A. C.; Levy, M. *Nucleic Acid Therapeutics* **2018**, *28* (3), 194–199.
- (4) Potty, A. S. R.; Kourentzi, K.; Fang, H.; Schuck, P.; Willson, R. C. *Int. J. Biol. Macromol.* **2011**, *48* (3), 392–397.
- (5) Haßel, S. K.; Mayer, G. *Molecular Diagnosis and Therapy* **2019**, *23* (3), 301–309.
- (6) Li, N.; Ebright, J. N.; Stovall, G. M.; Chen, X.; Nguyen, H. H.; Singh, A.; Syrett, A.; Ellington, A. D. *J. Proteome Res.* **2009**, *8* (5), 2438–2448.
- (7) Cox, J. C.; Ellington, A. D. *Bioorg. Med. Chem.* **2001**, *9*, 2525–2531.
- (8) Padlan, C. S.; Malashkevich, V. N.; Almo, S. C.; Levy, M.; Brenowitz, M.; Girvin, M. E. *Rna* **2014**, *20* (4), 447–461.
- (9) Dunn, M. R.; Jimenez, R. M.; Chaput, J. C. *Nature Reviews Chemistry* **2017**, *1* (10), 1–16.
- (10) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346* (6287), 818–822.

- (11) Tuerk, C.; Gold, L. *Science* **1990**, *249* (4968), 505–510.
- (12) Kirby, R.; Cho, E. J.; Gehrke, B.; Bayer, T.; Park, Y. S.; Neikirk, D. P.; McDevitt, J. T.; Ellington, A. D. *Anal. Chem.* **2004**, *76* (14), 4066–4075.
- (13) Mihai, I.; Vezeanu, A.; Polonschii, C.; David, S.; Gaspar, S.; Bucur, B.; Blaszykowski, C.; Sheikh, S.; Thompson, M.; Vasilescu, A. *Analytical Methods* **2014**, *6* (19), 7646–7654.
- (14) Mihai, I.; Vezeanu, A.; Polonschii, C.; Albu, C.; Radu, G. L.; Vasilescu, A. *Sensors and Actuators, B: Chemical* **2015**, *206*, 198–204.
- (15) Vasilescu, A.; Gaspar, S.; Mihai, I.; Tache, A.; Litescu, S. C. *Analyst* **2013**, *138* (12), 3530–3537.
- (16) Vasilescu, A.; Purcarea, C.; Popa, E.; Zamfir, M.; Mihai, I.; Litescu, S.; David, S.; Gaspar, S.; Gheorghiu, M.; Jean-Louis Marty. *Biosens. Bioelectron.* **2016**, *83*, 353–360.
- (17) Cho, E. J.; Collett, J. R.; Szafranska, A. E.; Ellington, A. D. *Anal. Chim. Acta* **2006**, *564* (1), 82–90.
- (18) Hybarger, G.; Bynum, J.; Williams, R. F.; Valdes, J. J.; Chambers, J. P. *Anal. Bioanal. Chem.* **2006**, *384* (1), 191–198.
- (19) Shamsipur, M.; Farzin, L.; Tabrizi, M. A. *Microchimica Acta* **2016**, *183* (10), 2733–2743.
- (20) Zuo, L.; Qin, G.; Lan, Y.; Wei, Y.; Dong, C. *Sensors and Actuators, B: Chemical* **2019**, *289*, 100–105.
- (21) Bayramoglu, G.; Ozalp, V. C.; Yilmaz, M.; Guler, U.; Salih, B.; Arica, M. Y. *Microporous Mesoporous Mater.* **2015**, *207*, 95–104.
- (22) Iacobucci, D.; Posavac, S. S.; Kardes, F. R.; Schneider, M. J.; Popovich, D. L. *Journal of Consumer Psychology* **2015**, *25* (4), 690–704.
- (23) DeCoster, J.; Gallucci, M.; Iselin, A.-M. R. *Journal of Experimental Psychopathology* **2011**, *2* (2), 197–209.
- (24) Cowperthwaite, M. C.; Ellington, A. D. *Journal of Molecular Evolution* **2008**, *67* (1), 95–102.
- (25) Zuker, M. *Nucleic Acids Res.* **2003**, *31* (11), 3406–3415.