

# Synthesis of Random DNA Libraries for In Vitro Selection and Analysis of Their Nucleotide Composition

A. S. Davydova<sup>a, 1</sup>, O. A. Krasheninina<sup>a</sup>, A. E. Tupikin<sup>a</sup>, M. R. Kabilov<sup>a</sup>,  
A. G. Venyaminova<sup>a</sup>, and M. A. Vorobyeva<sup>a</sup>

<sup>a</sup>*Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, 630090 Russia*

Received June 13, 2019; revised July 16, 2019; accepted July 22, 2019

**Abstract**—Aptamers are DNA or RNA molecules that specifically bind certain targets with high affinity due to the formation of unique spatial structures. Aptamers with specified properties could be obtained from combinatorial libraries of nucleic acids by in vitro selection procedure. Here, we optimized conditions of chemical synthesis of DNA library on the automated ASM-800 DNA/RNA synthesizer. Enzymatic hydrolysis and high-throughput sequencing methods were used to analyze compositions of synthesized combinatorial libraries. The DNA library generated under the chosen conditions is characterized by the sufficiently uniform randomization and can be applied for the in vitro selection.

**Keywords:** SELEX, combinatorial libraries of nucleic acids, solid-phase synthesis, high throughput sequencing

**DOI:** 10.1134/S1068162019060141

## INTRODUCTION

Currently, synthetic combinatorial libraries of nucleic acids are widely demanded for in vitro selection of nucleic acid aptamers. Aptamers are relatively short DNA or RNA fragments that bind non-covalently to specific target molecules due to the combination of unique nucleotide sequence and characteristic spatial structure. High affinity and specificity of target binding make nucleic acid aptamers a prospective alternative to monoclonal antibodies for the development of tools for targeted therapy and molecular diagnostics [1]. As a rule, synthetic combinatorial DNA or RNA libraries for aptamer selection consist of a 20–50 nt random region flanked by two constant primer-binding sites for reverse transcription and PCR amplification during the selection procedure [2]. The generation of DNA or RNA library starts from the automated solid-phase chemical synthesis of DNA templates using DNA/RNA synthesizer. Nowadays, several oligonucleotide synthesis companies provide commercial combinatorial libraries of nucleic acids for aptamer selection (see, e.g., Integrated DNA Technologies <https://eu.idtdna.com/site>). However, a considerable number of researchers still prefer homemade synthetic libraries. Uniformly randomized libraries are most commonly used for aptamer selection by a classic SELEX procedure. The term “uniformly randomized” means that each of the four nucleotides is represented at each position of the random region with approximately the same probability.

This type of random region provides the maximal diversity of nucleotide sequences and a high likelihood of successful selection of high-affinity aptamer [3, 4]. In the most protocols for the design and synthesis of combinatorial DNA libraries, authors strongly recommend optimization of chemical synthesis conditions for homemade libraries as a prerequisite for successful in vitro selection (see, e.g., [5]). In this context, the tuning of optimal monomer ratio is of foremost importance, to provide equally probable incorporation of each nucleotide during the synthesis of the random region. The goals of a present study were to optimize synthesis conditions of the DNA library on the automated DNA/RNA synthesizer ASM-800 and to analyze the uniformity of nucleotide composition for synthesized combinatorial libraries. The obtained DNA library could be further used both for the direct selection of DNA aptamers and as a template for subsequent transcription of RNA library.

## RESULTS AND DISCUSSION

**Optimization of synthesis conditions of random DNA libraries.** For this study, we used 87 nt DNA library 5'-GCCTGTTGTGAGCCTCCTGTCGAAN<sub>40</sub>-TTGAGCGTTTATTCTTGTCTCCC-3' with 40 nt random region flanked by two constant primer binding sites necessary for an enzymatic amplification during in vitro selection (PCR for DNA aptamers or in vitro RNA transcription and reverse transcription for RNA aptamers). We designed the sequences of constant sites by analogy with Fitzwater et al. [6] in a way that minimizes a formation of primer-dimers during

<sup>1</sup> Corresponding author: phone: +7 (383) 363-51-29; fax: +7 (383) 363-51-53; e-mail: [anna.davydova@niboch.nsc.ru](mailto:anna.davydova@niboch.nsc.ru).

PCR amplification, excludes complementary interactions between constant regions, and reduces the likelihood of abortive transcription initiation for the synthesis of combinatorial RNA libraries. A mixture of four standard fully protected DNA phosphoramidites was used for the synthesis of the random region of single stranded DNA library. It is important to emphasize that all four nucleotide monomers possess different coupling reactivity. Therefore, the use of an equimolar mixture of DNA monomers inevitably leads to unequal nucleotide composition [5]. Several variants of molar ratios of monomers for synthesis of uniformly randomized DNA libraries are given in the literature. However, they are not universal and would not be equally effective for automated synthesizers provided by different manufacturers. Taking into account published data on the coupling reactivity of commercial phosphoramidites together with our laboratory's experience in the field of oligonucleotide synthesis on automated ASM synthesizers, we chose two molar ratios: (1) A : C : G : T = 1.1 : 1 : 1.3 : 1 (library DL87-1); (2) A : C : G : T = 1.2 : 1 : 1.2 : 1 (library DL87-2).

**Determination of nucleotide composition uniformity for DNA libraries.** To determine the nucleotide composition of synthesized DNA libraries, we used complete enzymatic hydrolysis by phosphodiesterase I from *Crotalus adamanteus* venom followed by reversed-phase HPLC analysis of reaction products (Fig. 1). Reference values of retention times and spectral ratios for nucleoside markers were used to assign peaks of individual nucleosides [4].

After the chromatography data analysis with a MultiChrom program package, we estimated peak areas for each of four nucleosides and calculated their fractions in DNA library (see Table 1).

The fraction of each nucleoside in the DNA library was calculated using the equation  $\omega_n = (87\omega_0 - n)/40$ , where  $\omega_0$  is a total fraction of the particular nucleoside in the library (obtained from the chromatogram), and  $n$  is a number of its residues in constant regions of the library. The obtained values for all nucleosides are represented in Table 1. According to chromatography data analysis, DL87-2 library possesses more uniform nucleotide distribution of the random region, while the fractions of each nucleoside deviate more or less from the "ideal" value of 0.25.

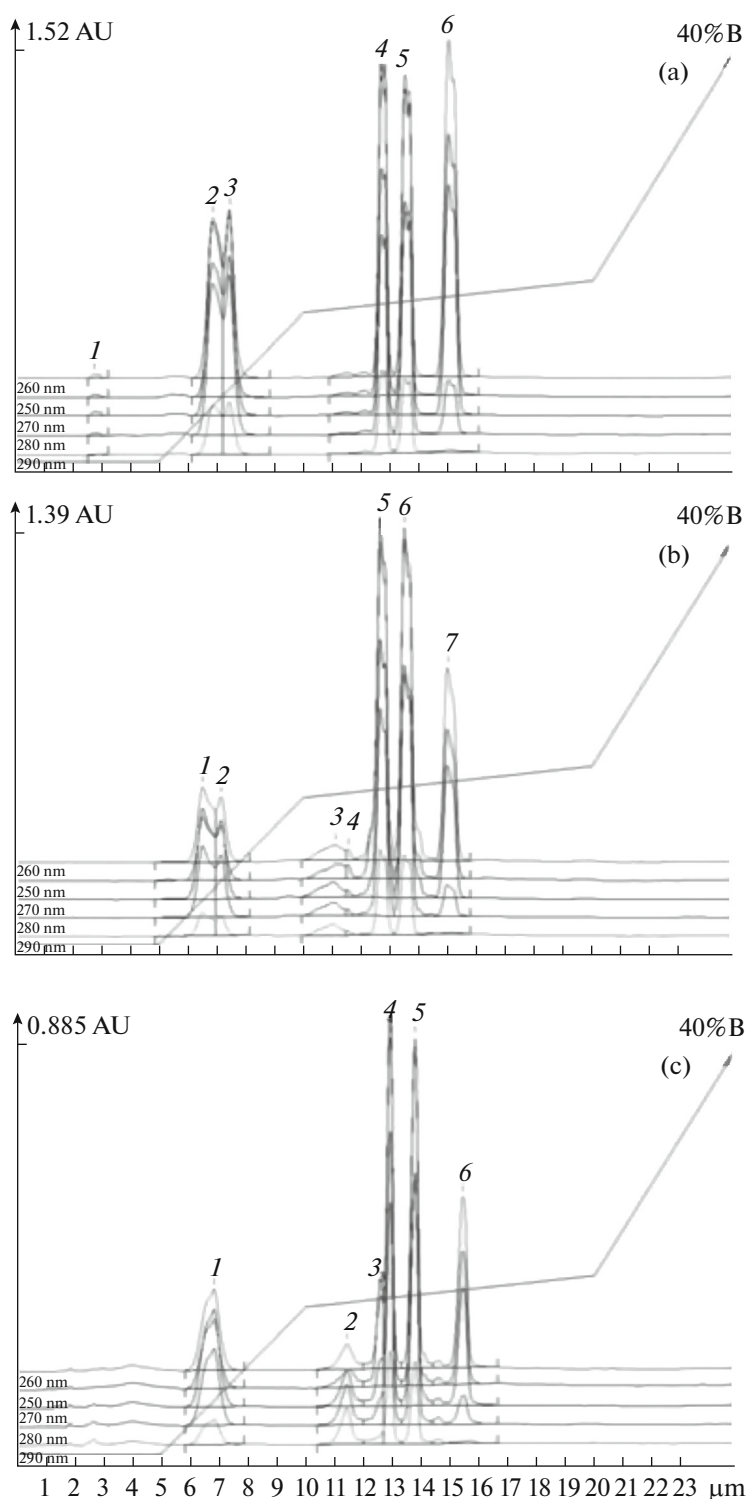
Of note, the summarized fractions of purine and pyrimidine nucleosides are almost equal, which can provide a sufficient diversity of structures for selection. We assume that insignificant deviation from uniform nucleotide distribution of the random region in the initial library should not reduce the probability of selection of high-affinity aptamers.

**Sequencing of a random DNA library.** For more in-depth analysis of the DL87-2 library, we performed its high-throughput sequencing using the MiSeq platform (Illumina). Chemically synthesized single-stranded DNA library was converted to a double-

stranded form using Klenow fragment (exo-). Then, adaptor sequences were introduced into the double-stranded DNA library by PCR amplification with specific primers. The library was sequenced on the MiSeq platform using the  $2 \times 300$  bp paired-ends sequencing kit (Illumina) at the Genomics Core Facility (ICBFM SB RAS, Novosibirsk, Russia). Data analysis for a sense strand of the DNA library revealed the following values of nucleoside fractions for the random region: 0.23 for A, 0.33 for C, 0.21 for G and 0.23 for T (see Table 2). Herein, at each particular position of the random region, the nucleotide ratio remained almost the same (Fig. 2).

Hence, the results of the sequencing of the initial DNA library have differed to some extent from those of the chromatographic analysis of complete enzymatic hydrolysis. According to literature data, the synthesis of complementary DNA strand by Klenow fragment should not significantly alter the nucleotide composition of the DNA library [6]. A possible reason for the observed inconsistency could be an unequal amplification of sequences with different GC content ("PCR bias") [8] during the PCR step of the sample preparation step before sequencing. Indeed, several publications reported a shift of nucleoside composition toward pyrimidine enrichment during the selection of RNA aptamers [9, 10], regardless of the type of the target. Despite restricted PCR application (4 cycles), high-throughput sequencing revealed 6% portion increase for C and 3% for T, which proves our supposition. Thereby, in our case the results of high-throughput sequencing didn't fully represent the nucleotide composition of the initial random DNA library. To overcome this problem, one might apply an alternative method of direct ligation of adaptors to the DNA library for further high-throughput sequencing. However, this approach also has its limitations since the specificities of the primary structure of individual sequences within the library can cause the formation of concatemers and adaptor dimers. This, in turn, might decrease the efficiency of enzymatic ligation. Of note, the shift of nucleoside distribution represents a problem only for the analysis of the initial library for in vitro selection. This does not influence the successful application of high-throughput sequencing for analysis of enriched libraries. During in vitro selection, PCR is a necessary step for the amplification of an enriched nucleic acid library at every round. Thus, a shift of nucleotide composition toward more effectively amplified sequences becomes quite inevitable. In this case, insignificant disturbance of the nucleotide composition of libraries due to the sequencing step does not contribute significantly to final results.

In summary, we performed the chemical synthesis of the initial single-stranded DNA library using the automated synthesizer ASM-800 and analyzed its nucleotide composition by two different methods. The optimization of synthetic conditions allowed obtaining libraries with a uniform nucleotide composition of



**Fig. 1.** Reversed-phase HPLC for products of enzymatic hydrolysis of DNA libraries. (a) Chromatographic analysis of control equimolar mixture of four nucleoside markers. (b) Chromatographic analysis of enzymatic hydrolysis' products for DNA library DL87-1. C. Chromatographic analysis of enzymatic hydrolysis' products for DNA library DL87-2. Mobile phase 0.02 M TEAAc, pH 7.0 (aq.)/ 0.02 M TEAAc, pH 7.0 (50% acetonitrile), Prontosil-120-5-C18 AQ column. All peaks are numbered; spectral ratios are indicated in Table 1.

**Table 1.** Spectral ratios and fractions of each of four nucleosides in synthesized DNA libraries

	No. of peak	A <sub>250</sub> /A <sub>260</sub>	A <sub>280</sub> /A <sub>260</sub>	A <sub>250</sub> /A <sub>260</sub>	A <sub>280</sub> /A <sub>260</sub>	Nucleoside	$\omega_0$	$n$	$\omega(N)$
Mix of nucleoside markers	2	0.84	0.93	0.83	0.97	C	—	—	—
	3	0.84	0.94						
	4	1.14	0.65	1.16	0.68	G	—	—	—
	5	0.65	0.72	0.65	0.70	T	—	—	—
	6	0.78	0.17	0.79	0.15	A	—	—	—
DL87-1	1	0.86	0.93	0.83	0.97	C	0.26	13	0.23
	2	0.86	0.93						
	5	1.14	0.65	1.16	0.68	G	0.26	10	0.32
	6	0.65	0.72	0.65	0.70	T	0.34	18	0.28
	7	0.77	0.17	0.79	0.15	A	0.14	5	0.17
DL87-2	1	0.86	0.94	0.83	0.97	C	0.27	13	0.26
	4	1.14	0.66	1.16	0.68	G	0.24	10	0.27
	5	0.65	0.73	0.65	0.70	T	0.33	18	0.27
	6	0.78	0.22	0.79	0.15	A	0.15	5	0.20

**Table 2.** Nucleotide composition of random regions of synthesized DNA libraries DL87-1 and DL87-2 for antisense (–) and sense (+) strands

Library	Synthesis ratio for A : C : G : T	Strand	$\omega(A)$	$\omega(C)$	$\omega(G)$	$\omega(T)$	Evaluation method
DL87-1	1.1 : 1 : 1.3 : 1	–	0.17	0.23	0.32	0.28	Chromatography
		+	0.28	0.32	0.23	0.17	Calculation
DL87-2	1.2 : 1 : 1.2 : 1	–	0.20	0.26	0.27	0.27	Chromatography
		+	0.27	0.27	0.26	0.20	Calculation
		+	0.23	0.33	0.21	0.23	Sequencing

the random region. As previously stated, chemically synthesized single-stranded DNA library could be used directly for DNA aptamer selection, and also

could serve as a template for RNA library synthesis for the selection of RNA aptamers. The obtained DNA library DL87-2 was further successfully used as a tem-

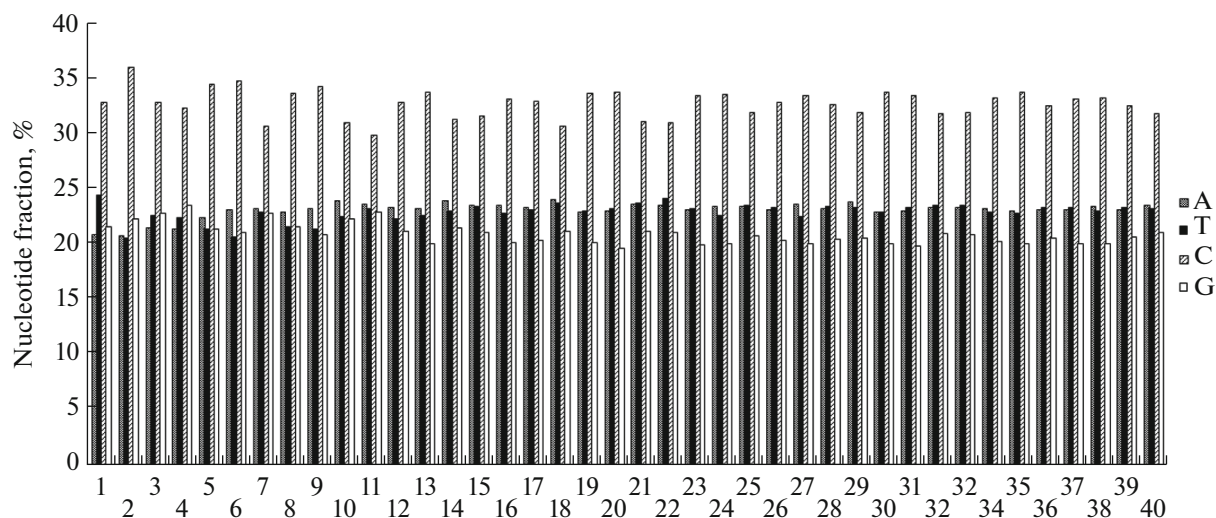
**Fig. 2.** Nucleotide distribution at positions 1 to 40 of the random region according to data analysis of high-throughput sequencing on the MiSeq platform (Illumina).

plate for enzymatic synthesis of a combinatorial library of 2'-F-RNA molecules for in vitro selection of 2'-F-modified RNA aptamers specific to photoprotein obelin [11] and human hemoglobin and glycated hemoglobin [12]. Among selected RNA aptamers, there were both sequences with a uniform nucleotide composition which form stem-loop secondary structures, and G-rich sequences giving G-quadruplexes. This points to the fact that the nucleotide composition of the random region within the synthesized DNA library provides a high diversity of nucleotide sequences sufficient for the selection of aptamers of different structure types.

## EXPERIMENTAL

**Synthesis and purification of randomized single-stranded DNA libraries.** Synthesis of oligodeoxyribonucleotides was performed using the automated synthesizer ASM-800 (Biosset, Novosibirsk, Russia) in a 0.4  $\mu$ mole scale by the solid-phase phosphoramidite synthesis protocols optimized for the particular instrument. 5-(Ethylthio)-1H-tetrazole (Sigma-Aldrich, USA) was used for the condensation step of the automated synthesis. Single-stranded DNA library 5'-GCCTGTTGTGAGCCTCCTGTCTCGAAN<sub>40</sub>TTGAGCGTTTATTCTTGTCTCCC-3' (N, any nucleotide) consisted of 40 nt random region flanked by two constant primer binding sequences. A mixture of four standard commercially available fully protected deoxynucleotide phosphoramidites (GlenResearch, USA) was used for random region synthesis. Molar ratios of four monomers are given in the "Results and discussion" section. All oligonucleotides were deprotected according to the manufacturer's recommendation [13]. DNA libraries were purified by denaturing PAGE (12% acrylamide, 8 M urea), eluted from the gel and concentrated using Amicon Ultra-0.5 10K (Millipore, USA).

**Enzymatic hydrolysis of DNA libraries.** Purified DNA library (500 pmole) was incubated with 0.01 u of phosphodiesterase I from *Crotalus adamanteus* venom (Sigma-Aldrich, USA) at 37°C for 16 h in total volume of 50  $\mu$ L of the reaction mixture. Then 10 u of thermosensitive alkaline phosphatase FastAP (Thermo Fisher Scientific, USA) was added, and the reaction mixture was incubated for another 30 min at 37°C. Afterward, the reaction mixture was incubated at 75°C for 10 min to inactivate alkaline phosphatase. The reaction mixture was analyzed by reversed-phase HPLC on the Millichrome chromatograph (EcoNova Ltd., Russia), using the ProntoSil-120-5-C18 AQ column and 0–50% concentration gradient of acetonitrile in 0.02 M TEAAc (pH 7.0) for 25 min at 100  $\mu$ L/min.

**Synthesis of double-stranded random DNA libraries.** Double-stranded random DNA library was synthesized by the DNA primer elongation reaction by the Klenow fragment (exo–) of *E. coli* DNA polymerase I

(Thermo Fisher Scientific, USA). The reaction mixture (100  $\mu$ L) containing 1 nmole of DNA library and 2 nmoles of forward primer in 10 M Tris-HCl (pH 8.0) and 10 mM MgCl<sub>2</sub> was incubated at 90°C for 5 min and then cooled down to room temperature for 15 min. Annealed library-primer duplex was then added to the reaction mixture containing 50 mM Tris-HCl (pH 8.0), 10 mM MgCl<sub>2</sub>, 50 mM NaCl, 0.5 mM dATP, 0.5 mM dGTP, 0.5 mM dCTP, 0.5 mM TTP and 50 u Klenow fragment (exo–). The mixture was incubated for 30 min at 37°C and then heated for 10 min at 75°C to inactivate the enzyme. The reaction product was purified using the MinElute Reaction Cleanup Kit (Qiagen, Germany).

**Sequencing of DNA libraries and data analysis.** To synthesize the DNA library for sequencing, 1  $\mu$ g of double-stranded DNA library was re-amplified with the specific primers containing adapters and barcodes for high-throughput Illumina sequencing. Restricted amplification (4 cycles) was performed using reagents of NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, UK). The obtained DNA library was sequenced on the MiSeq platform (Illumina) using 2  $\times$  300 bp paired-ends sequencing kit (Illumina) in Genomics core facility (ICBFM SB RAS, Novosibirsk, Russia).

Bioinformatic analysis of raw sequences was performed using CLC GW 10.0 (Qiagen) and included trimming of adaptor sequences from 3'-end and primers from 5'-end of paired reads, read quality filtering and sequential overlapping. The nucleotide frequencies in the 40 nt random region was calculated using the Python3 script.

## FUNDING

The work was supported by the Russian Scientific Foundation, project no. 16-14-10296 and by Russian State funded budget project No. AAAA-A17-117020210021-7 to ICBFM SB RAS.

## COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain any studies involving animals or human participants performed by any of the authors.

## Conflict of Interests

Authors declare that they have no conflicts of interests.

## REFERENCES

1. Vorobyeva, M., Vorobjev, P., and Venyaminova, A., *Molecules*, 2016, vol. 21, p. 1613.
2. Cowperthwaite, M.C. and Ellington, A.D., *J. Mol. Evol.*, 2008, vol. 67, pp. 95–102.
3. Blind, M. and Blank, M., *Mol. Ther. Nucleic Acids*, 2015, vol. 4, e223.

4. Takahashi, M., Wu, X., Ho, M., Chomchan, P., Rossi, J.J., Burnett, J.C., and Zhou, J., *Sci. Rep.*, vol. 6, p. 33697.
5. Hall, B., Micheletti, J.M., Satya, P., Ogle, K., Pollard, J., and Ellington, A.D., *Curr. Protoc. Nucleic Acid Chem.*, 2009, chapter 9, unit 9.2.
6. Fitzwater, T. and Polisky, B., *Methods Enzymol.*, 1996, vol. 267, pp. 275–301.
7. Dawson, R., Elliott, D., Elliott, W., and Jones, K., *Data for Biochemical Research*, 3rd ed., Oxford: Clarendon, 1986.
8. Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A., *Genome Biol.*, 2011, vol. 12, p. R18.
9. Thiel, W.H., Bair, T., Wyatt, ThielK., Dassie, J.P., Rockey, W.M., Howell, C.A., Liu, X.Y., Dupuy, A.J., Huang, L., Owczarzy, R., Behlke, M.A., McNamara, J.O., and Giangrande, P.H., *Nucleic Acid Ther.*, 2011, vol. 21, pp. 253–263.
10. Takahashi, M., Wu, X., Ho, M., Chomchan, P., Rossi, J.J., Burnett, J.C., and Zhou, J., *Sci. Rep.*, 2016, vol. 6, p. 33697.
11. Davydova, A.S., Vorob'eva, M.A., Krasitskaya, V.V., Tupikin, A.E., Vorob'ev, P.E., Kabilov, M.R., Frank, L.A., and Ven'yaminova, A.G., *Acta Naturae*, 2017, vol. 9, p. 15.
12. Davydova, A., Vorobyeva, M., Bashmakova, E., Vorobjev, P., Krasheninina, O., Tupikin, A., Kabilov, M., Krasitskaya, V., Frank, L., and Ven'yaminova, A., *Anal. Biochem.*, 2019, vol. 570, pp. 43–50.
13. www.glenresearch.com, section deprotection guide.