

# Orthogonal DNA barcodes enable the high-throughput screening of sequence-defined polymers

Donatien de Rochambeau<sup>1†</sup>, Serhii Hirka<sup>1†</sup>, Daniel Saliba<sup>1</sup>, Shaun Anderson<sup>1</sup>, Violeta Toader<sup>1</sup>, Michael D. Dore<sup>1</sup>, Felix J. Rizzuto<sup>1</sup>, Maureen McKeague<sup>1,2\*</sup>, and Hanadi F. Sleiman<sup>1\*</sup>

<sup>1</sup>Department of Chemistry, McGill University, 801 Sherbrooke St W, Montréal, QC H3A 08B, Canada

<sup>2</sup>Department of Pharmacology and Therapeutics, McGill University, 3655 Prom. Sir William Osler, Montreal, Quebec, H3G 1Y6, Canada

<sup>†</sup> These authors contributed equally.

---

**ABSTRACT:** Polymers with sequence definition allow access to programmable morphologies and applications, but directly correlating polymer structure to function currently requires case-by-case analysis: high throughput methods that identify promising species from entire chemical families are required. Here, we show that the discovery of effective protein target-recognition molecules can be achieved using DNA-encoded libraries of chemically diverse sequence-defined oligomers, generated on an automated DNA synthesizer. These structures are ALENOMERs – Aptamer-Like ENcoded OligoMERs – that are read and sequenced using a DNA code that branches from, and corresponds to, the target-binding oligomer. By incorporating nucleosidic and non-nucleosidic components into alenomers at specific locations, we unlock new supramolecular interactions for biomolecule binding, and directly correlate their effectiveness at each site. Our alenomer library screening removes the low throughput bottleneck of analyzing individual sequence-defined polymers, improving the binding efficacies of natural systems and enabling wide chemical spaces to be sampled for biomolecule sensing, therapy, and diagnostics.

---

Sequence-defined polymers (SDPs) made of both natural and unnatural building blocks have the potential to program unique morphologies and functions into biopolymer derivatives.<sup>1, 2</sup> However, the emergence of applications that require SDPs is slowed by the need to analyse individual polymers one by one. Large combinatorial libraries help overcome this issue, as functional space can be screened in high-throughput experiments.<sup>3</sup> Initially developed for small molecules,<sup>4-6</sup> combinatorial libraries have been expanded to artificial oligocarbamates,<sup>7</sup> peptoids,<sup>8</sup> and thiolactone precision polymers<sup>9</sup> that have found applications as protein ligands,<sup>7, 10</sup> antimicrobial agents<sup>8</sup> and drug solubilizers<sup>9</sup>. The selection and identification of a single member

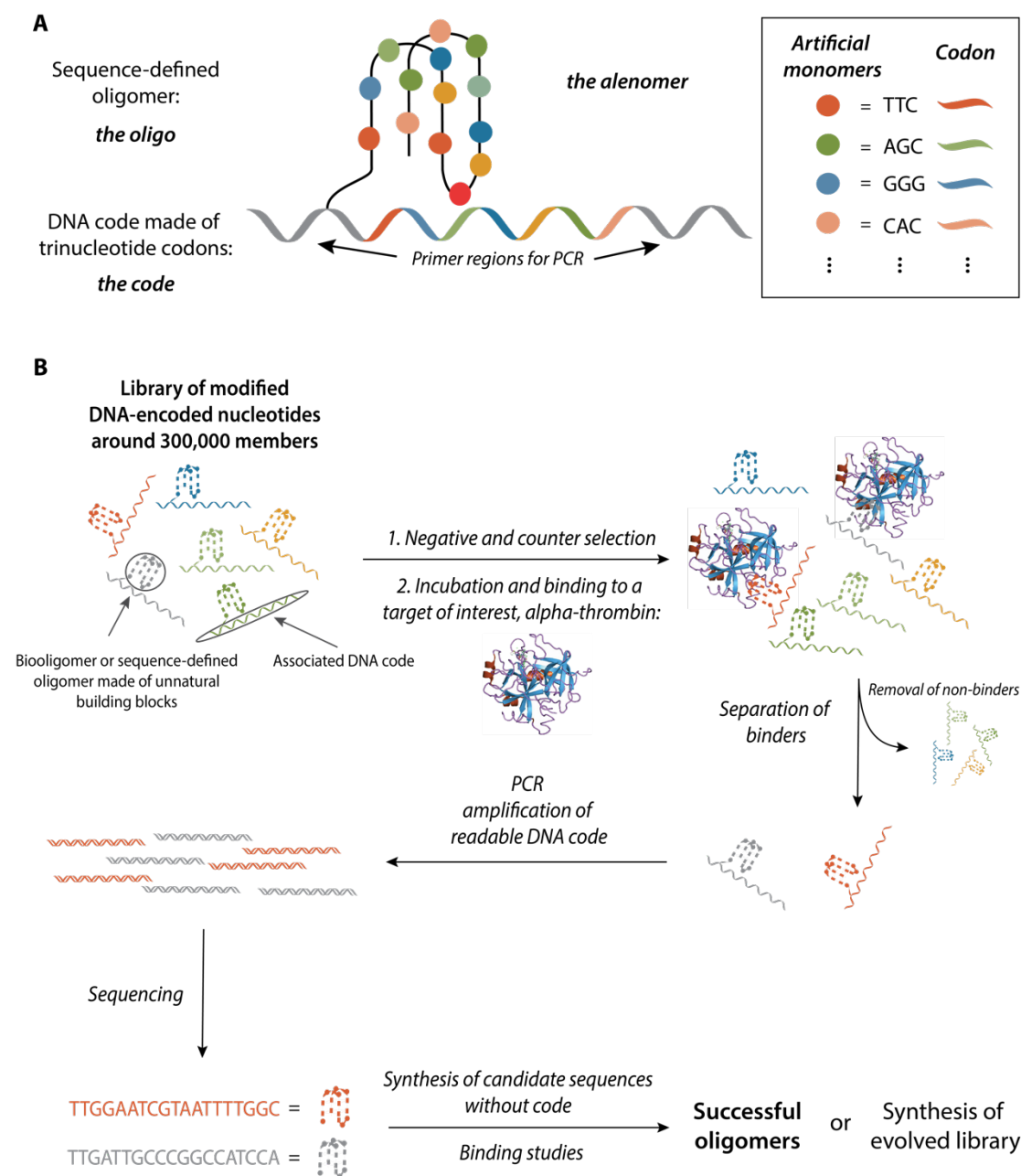
of a SDP library is not trivial, however, and is complicated by the dearth of methods to sequence artificial SDPs.<sup>11</sup>

An attractive strategy to facilitate the identification of effective SDPs from large libraries would be to assign a DNA barcode to each library member. Such DNA-encoded libraries (DELs) have already greatly facilitated high-throughput screening for small molecules, and they are now a mainstay in drug discovery.<sup>12</sup> DELs encrypt each small molecule library member with a DNA code<sup>13</sup> that can be amplified and sequenced from very small concentrations ( $<10^{-15}$  nM) using the polymerase chain reaction (PCR);<sup>14-18</sup> DEL synthesis<sup>16</sup> typically requires organic reactions that are orthogonal to DNA-ligation chemistry, and may not provide the high coupling yields necessary for making sequence-defined polymers. Solid-phase phosphoramidite synthesis, on the other hand, can afford high-purity DNA strands<sup>19</sup> up to 200 nucleotides long, and is even used to produce FDA-approved oligonucleotide therapeutics.<sup>20</sup> Our group<sup>21-23</sup> and others<sup>24-29</sup> have demonstrated the use of automated solid-phase phosphoramidite synthesis to generate artificial sequence-defined oligo(phosphodiester)s in high yields and with large chemical diversity. Solid-phase synthesis can thus in principle be used to both construct the SDP as well as its DNA barcode. This method would significantly expand the chemical space of DELs, enabling applications of sequence-defined polymers in protein binding, therapeutics, diagnostics, and enzyme-mimics.<sup>1</sup>

We hypothesized that such a DEL would effectively select and identify new aptamer-like SDPs. Aptamers are nucleoside sequences that recognize and bind to target molecules and are largely limited to the natural nucleobase code. Introducing artificial SDP segments into aptamer constructs would improve their potential for high affinity binding by expanding the chemical space they can explore.<sup>30</sup> Efforts to impart chemical diversity into nucleic acids aptamers – by introducing chemically modified DNA and RNA nucleobases, or adding single nucleotide modifications into aptamers post-selection, for example – have led to improvements in affinity and *in vivo* stability, thus expanding their applications in health and medicine.<sup>31-38</sup> Almost all these methods require compatibility of the chemically modified nucleotides with polymerases or ligases, and therefore are relatively conservative in their diversity. A high-throughput strategy which incorporates non-nucleosidic modifications that are not necessarily recognized by enzymes is attractive, as it would significantly enlarge the structural and functional diversity of aptamers. Aptamer-Like DNA-ENcoded oligoMER (which we herein term *Alenomer*) libraries can thus

solve bottlenecks in both sequence-defined polymers and functional aptamer discovery, opening new potential for biomedical and materials applications.

We developed DNA-encoded libraries of chemically diverse sequence-defined oligomers to achieve the high-throughput discovery of effective target-recognition compounds. The specific nature and position of each non-nucleosidic unit is encoded by a three-DNA base codon residing on the covalently linked DNA arm (Figure 1A). We used the split-and-pool strategy<sup>3, 39</sup> to synthesize a nearly 300,000-member library of alenomers, specifically targeting G-quadruplex aptamer designs<sup>40</sup>. Libraries thus formed were subject to biomolecule selection, separation, and code amplification (Figure 1B). Our non-nucleosidic building blocks impart diverse molecular recognition interactions – hydrogen-bonding,  $\pi$ - $\pi$ , electrostatic interactions and the hydrophobic effect – to aptamer designs, translating into improved binding properties. Alenomers identified from next-generation sequencing of the DNA code showed improvements in binding affinity and serum stability, underpinning our strategy as an effective method to identify new and useful SDP sequences for biomolecule binding.



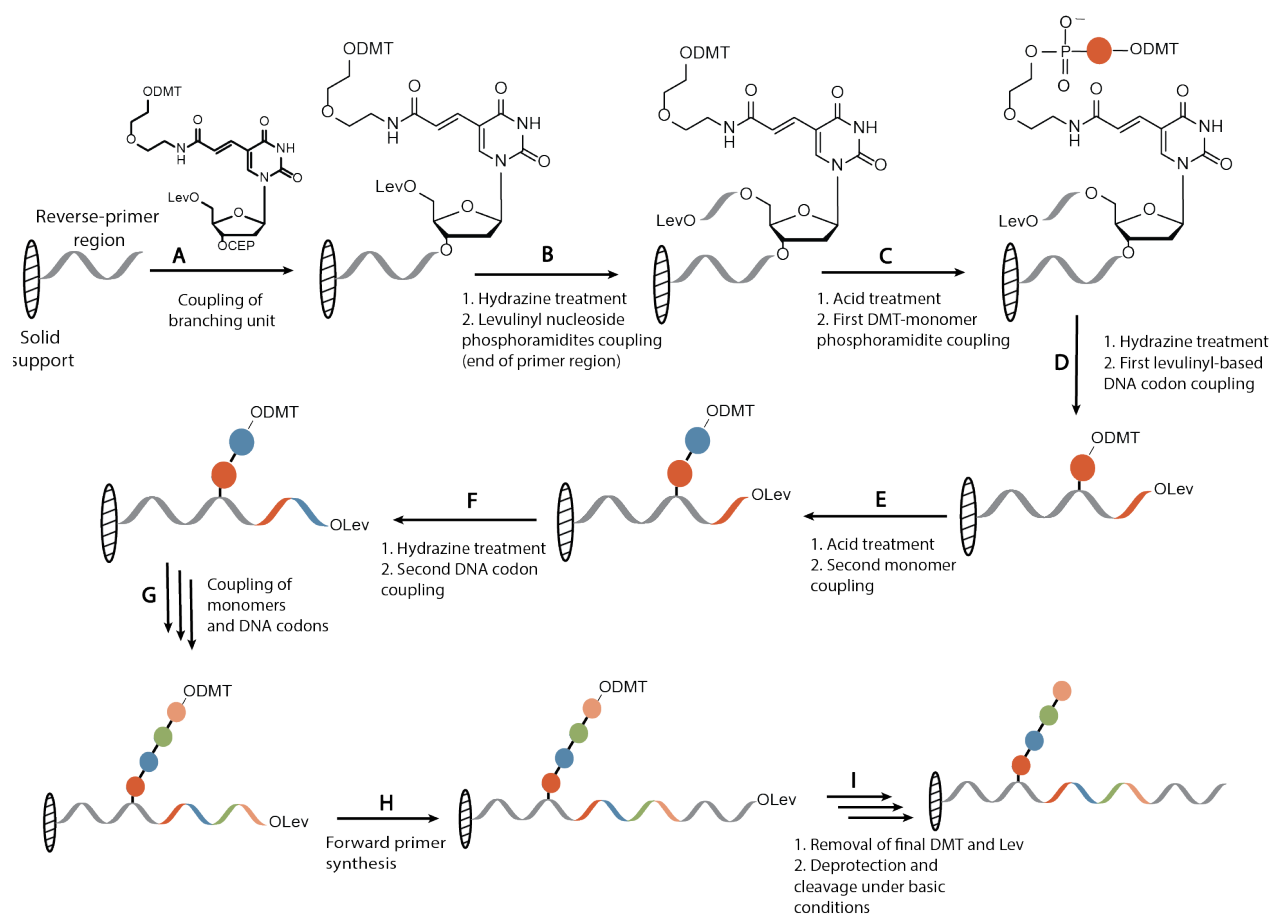
**Figure 1. General design of our alenomers, their incorporation into DNA libraries, and their selection and amplification to identify effective target-binding oligomers.** (A) Our alenomers are composed of a sequence-defined polymer with non-nucleosidic modifications (*the oligo*) that correspond to and branch from a DNA code (*the code*). (B) Large libraries of alenomers can be screened to bind to a target of interest. The non-nucleosidic chemical identity of each alenomer can be amplified by polymerase chain reaction (PCR) of the corresponding DNA code, enabling identification of promising protein-binders from among *ca.* 300,000 library members.

## Results and discussion

### Orthogonal synthesis of a branched DNA construct

The general design of our DNA-encoded non-nucleosidic oligomers relies on each monomer of the polymer backbone corresponding to a unique DNA code. Following PCR and sequencing, the amplified DNA code can then be correlated to a functional oligomer sequence, which reports on important modifications for activity or function. Our system is comprised of five main components to realise this goal: 1) A reverse primer region; 2) A branching junction point from which: 3) The functional oligomer sequence and 4) DNA codon sequence diverge; and 5) A forward primer region terminating the DNA code (Figure 1A).

The success of our strategy relies on the functional oligomer and the DNA code segments growing independently from a branched nucleotide junction, necessitating different and orthogonal protecting groups for solid-state synthesis; phosphoramidites with levulinyl (Lev) moieties<sup>41, 42</sup> were used to grow the DNA code, while phosphoramidites for the aptamer region were protected with dimethoxytrityl (DMT) moieties. These two functional groups are orthogonal (DMT is acid labile, Lev groups are cleaved with hydrazine), resistant to all reagents used in a commercial DNA synthesizer and are quickly and quantitatively removed. For the branch point, we used a deoxyuridine monomer modified with an acrylamido side chain at the C5 position (**BU**), which was synthesised in moderate yield from commercial starting materials in five steps (Scheme S1). This modification was selected to allow efficient DNA hybridization and can be recognised by some polymerases efficiently.<sup>35, 43</sup> Our system can be grown on a solid support using a commercial DNA synthesizer (Figure 2), using iterative steps of monomer growth followed by the corresponding synthesis of the code.



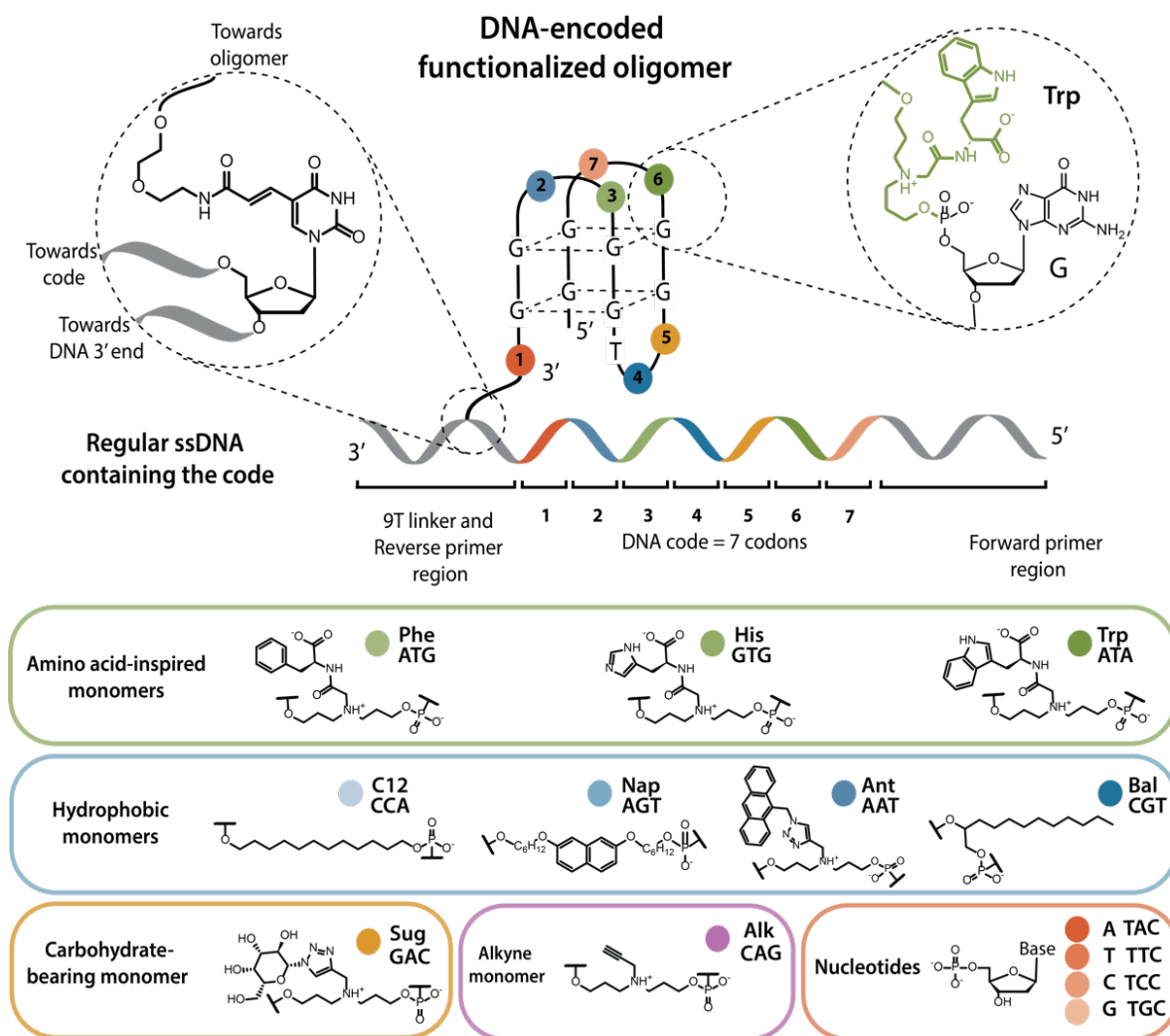
**Figure 2. Synthesis of alenomers using a commercial DNA synthesizer.** Our orthogonal branching strategy enables each monomer of the oligo to have a corresponding DNA code, which are added iteratively until the desired oligomer length and sequence is reached. Having optimized the position of the branch, the initial 12 nucleotides of the reverse primer region are added,<sup>44</sup> followed by the branching unit (step A) and the remaining four nucleotides of the reverse primer (step B). After deprotection of the DMT on the branching unit, a DMT-protected monomer is connected (step C). The Lev moiety on the branching unit is then cleaved and the corresponding DNA code segment added (step D). This process of DMT deprotection, oligomer monomer addition, Lev deprotection and DNA code addition is performed iteratively (steps E-G). Following full growth of the desired branched strand, the forward primer sequence of 11 nucleotides was added (step H) to enable downstream amplification and sequencing. The terminal orthogonal protecting groups were removed and the alenomers were cleaved from the solid support and deprotection under basic conditions (step I). Colours match each monomer of the synthetic oligomer to the orthogonally-synthesized DNA code.

## Inclusion of synthetic monomers

The use of the DNA code, which can be amplified and sequenced, enables the introduction of non-nucleosidic modifications into the sequence-defined oligomer. We thus developed a panel of nine synthetic non-nucleosidic phosphoramidites based on three synthetic platforms: directly from diols, using Cu<sup>I</sup> click chemistry, or from amides (Figure 3 and Scheme S2).<sup>21</sup>

Our aim in selecting specific functionalities was to merge the chemical diversity and function of all the major classes of biomolecules: monomers relevant to proteins (**Phe**, **His**, **Trp**), lipids (**C12**, **Bal**), carbohydrates (**Sug**), the common DNA monomers, as well as synthetic monomers = (**Ant**) and handles to add post-synthetic functionality (**Alk**). These modifications (Figure 3) present diverse intermolecular interactions –  $\pi$ -interactions, polar and hydrogen-bonding interactions, and the hydrophobic effect – that are absent, weaker, or simply different from purely-nucleoside structures. We confirmed the high yielding incorporation of the nine unnatural monomers (**Alk**, **Ant**, **Bal**, **C12**, **His**, **Phe**, **Nap**, **Sug**, **Trp**) at internal positions of a DNA strand and their compatibility with hydrazine (Figure S9).

We integrated our unnatural monomers, each encoded by a three-nucleotide DNA codon, into our branched DNA oligomer synthesis. Our SDPs were inspired by the G-quadruplex scaffold of the aptamer oligomer that binds to alpha thrombin,<sup>45</sup> due to its biological relevance and its well-characterized interaction with aptamers.<sup>46</sup> The G-quadruplex is a recurring motif in many aptamers and provides an ordered and thermodynamically stable scaffold.<sup>40</sup> Three model alenomers (**TBA1-3**) with varying numbers of modifications were synthesized and purified to confirm the utility and scalability of our platform (Table S1, Figures S8 and S22). To ensure that our branched design did not impede oligomer binding, the SDP structure of **TBA1** (a branched oligomer with the TBA aptamer attached to a DNA code, **Table S1**) was designed to mimic the natural DNA-only thrombin binding aptamer. We compared the function of the original thrombin binding aptamer (**TBA**) with **TBA1**. No significant impact on binding affinity to the target thrombin was observed by fluorescence anisotropy studies ( $K_d(\text{TBA}) = 11 \pm 8$  nM;  $K_d(\text{TBA1}) = 13 \pm 10$  nM, (Table S6). These results suggest that the DNA codes does not significantly inhibit binding, and that our alenomers are appropriate platforms from which oligomer libraries can be constructed.

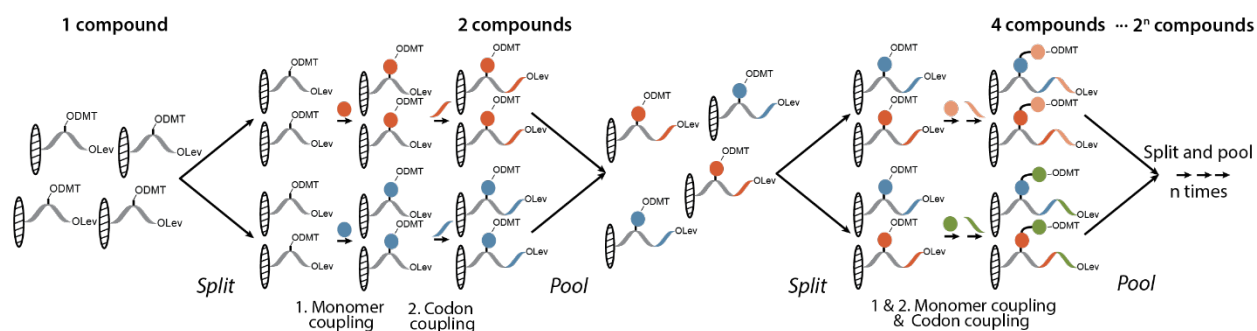


**Figure 3. Design of alenomers modelled after the G-quadruplex, made of a combination of natural and synthetic monomers.** Monomer names, structures and codons are shown in the bottom. The oligomer design is composed of the synthetic functional oligomer coupled to a DNA-based code through a branching unit. Primers (shown as grey ribbons) were included to enable enzymatic amplification (PCR) of the DNA code. A 3' linker of 9 thymidine units separates the DNA-encoded oligomer from the solid support and is followed by the first 12 bases of the reverse primer, the branching unit, and then the remaining 4 bases of the reverse primer. Extending from the branching unit, the functional branch of the oligomer is 4 (linker) + 7 (unnatural monomers) + 9 (unmodified positions) = 20 monomers in length, while the DNA code is 4 (end of reverse primer region) + 21 (7x3 nucleotides coding for the unnatural monomers) = 25 nucleotides long. The monomers used in this work include synthetic amino-acid like monomers (green), hydrophobic residues (blue), carbohydrate bearing units (yellow), a clickable-alkyne monomer (purple) as well as the four natural DNA nucleobases (orange).



## Synthesis of the alenomer library

A diverse library was synthesized so that the utility of the branched DNA approach could be evaluated for high-throughput functional screening and compatibility with next-generation sequencing. To maintain the secondary structure of the G-quadruplex scaffold, seven positions within the loop regions were modified with either nucleotides or unnatural monomers (6 different modifications). We applied a split-and-pool strategy<sup>39, 44</sup> to make an alenomer library (**LIB**) (Figure 4). When reaching position  $n$  of the sequence, the solid support was split into 6 aliquots. An unnatural monomer or nucleotide was added, followed by the attachment of the associated codon on the code branch. The solid supports were then pooled and mixed and split again for the next coupling. Addition of the G bases required to maintain the quadruplex backbone was executed without split and pool and thus also did not require DNA codon synthesis.



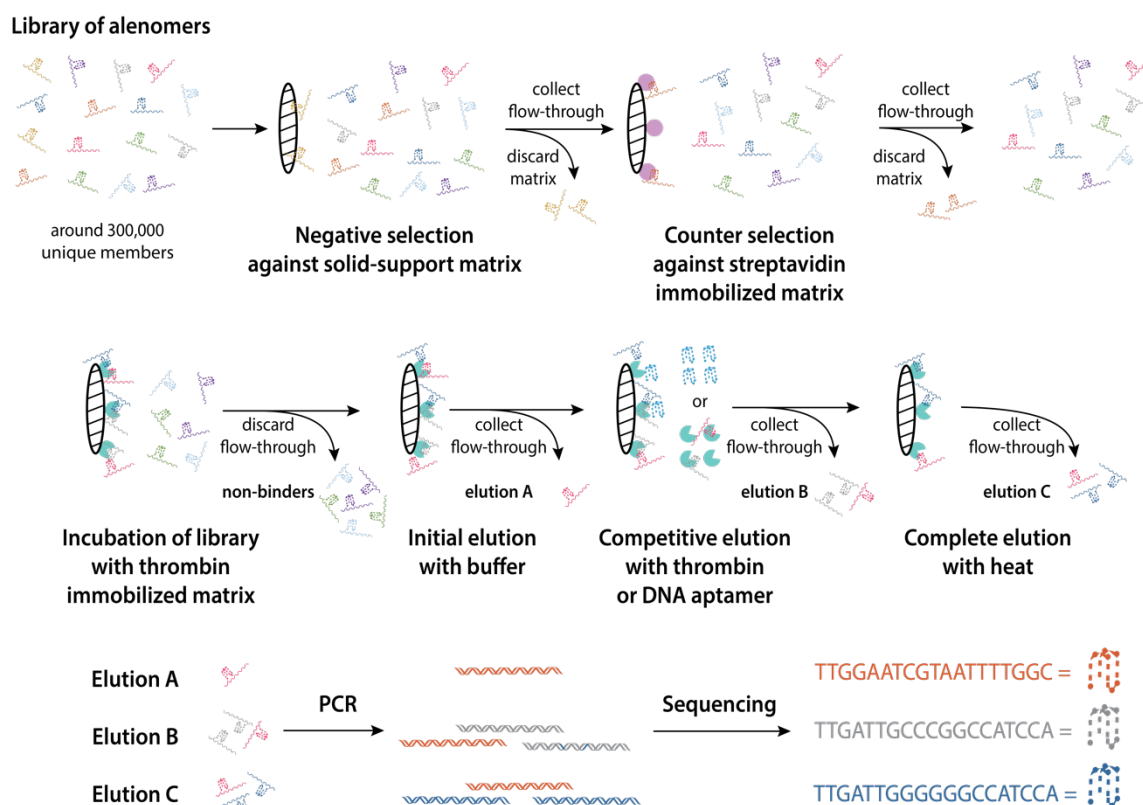
**Figure 4. Split and pool strategy for the synthesis of an alenomer library.** We show an example with 2 different monomers at each step for clarity. In this study, we couple 6 different monomers at each step yielding  $6^n$  compounds.

Libraries synthesized in this manner should be composed of  $6^7 = 279,936$  library members. We observed the presence of a relatively diffuse band in denaturing PAGE, at similar mobilities to the previously synthesized model alenomers (Figure S8), suggesting reasonable homogeneity in sequence length and size between all library components. Yields of synthesis of  $\sim 11\%$  were obtained (by gel image analysis). We isolated 2.5 nmols of the library using about  $1/6^{\text{th}}$  of the crude product obtained.

## Developing a high-throughput screening method for SDPs

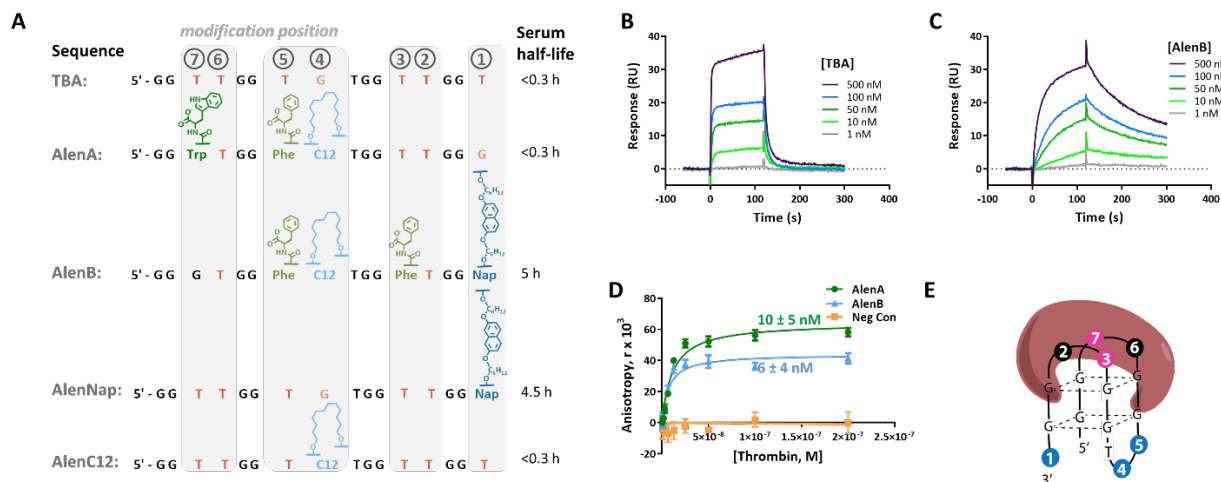
Functional nucleic acid polymers such as ribozymes, DNAzymes, and aptamers developed through synthetic protocols such as the Systematic Evolution of Ligands by Exponential enrichment (SELEX) require 10-15 rounds of selection and amplification.<sup>46</sup> As such, these protocols typically make use of nucleoside analogues that are compatible with enzymes for amplification or ligation. Our highly diverse alenomer library thus presents a first opportunity to screen for new functional oligomers containing a multitude of non-nucleosidic modifications.

Once the barcode design was validated by sequencing results (Table S10), the combinatorial library of alenomers was screened for high affinity binding to thrombin immobilized to a solid-support matrix. In a single round, alenomer candidates were collected using sequentially stronger elution steps ranging from gentle washes with buffer, to competition with DNA-based thrombin aptamers, to rigorous elution with high concentrations of the thrombin target itself (Figure 5). Due to the DNA barcode, the chemical diversity and identity of the oligomers collected in each fraction could be sequenced using next-generation sequencing. While typical selection experiments of new functional nucleic acids require many rounds and thus several months to identify potential binding candidates, this method could be performed in a single day. We repeated the same experiment an additional two times to confirm reproducibility; one of the experiments was performed with a fluorescently tagged library to monitor the fraction collection process (Figure S12).



**Figure 5. Stepwise screening principle of DNA-encoded oligomer libraries.** The strategy is inspired by Brenner and Lerner<sup>13</sup> who initially proposed the synthesis of DNA-encoded peptides on beads. The library is first subjected to negative and counter screens to remove candidates that non-specifically interact with the solid-support matrix or generic protein targets. Next, the library of oligomers is incubated with the target of interest and non-functional oligomers are removed with several washing steps. Functional oligomers are recovered in a stepwise fashion with increasingly stringent elution fractions. The DNA codes from each fraction from the screen are amplified and sequenced, revealing the chemical identity of functional sequences from the stepwise screen.

Sequencing results from each fraction confirmed that the rapid screening process enabled a 2000-7000-fold enrichment of the best oligomers in the library (see Table S11). Interestingly, the unmodified original thrombin aptamer (**TBA**) emerged as one of these enriched top candidates, confirming the validity of this approach. Two new alenomers, **AlenA** and **AlenB**, were identified in the top five most abundant sequences within each experiment and each competitive elution fraction. A list of the top 20 sequences for each experiment is available in a separate supplementary file. Figure 6A describes the identity of the monomers (nucleosidic or non-nucleosidic) within the full oligomer. Importantly, sequencing results from repeated experiments were similar – where the most abundant sequences were again identified with similar frequency.



**Figure 6. Alenomers identified by our high-throughput screen have improved binding and stability properties.** (A) The sequence name, composition, and serum stability of the three most popular alenomers identified in our screen along with comparison model TBA aptamer. Sequences are described 5' to 3'. Each position of the library that was identified via the DNA code is indicated in grey (1 through 7). The modifications are colour-coded according to Figure 3, with non-nucleosidic modifications shown and G and T representing typical DNA nucleobases. (B) SPR sensorgrams of increasing concentrations of the thrombin binding DNA aptamer (TBA) interacting with thrombin immobilized to a gold chip. (C) SPR sensorgrams of increasing concentrations of the alenomer **AlenB** interacting with thrombin immobilized to a gold chip. The double-referenced sensorgrams were fit to a 1:1 kinetic binding model. (D) Binding isotherms of alenomers **AlenA** and **AlenB** interacting with thrombin obtained using fluorescence anisotropy. A scrambled DNA sequence is included as a negative control ( $n = 3$ ). (E) An exemplary alenomer library member interacting with thrombin.

## High-content analysis of the oligomer modification landscape

Conventional aptamers or other SDPs can be synthesized with numerous modifications including lipids to enable cell uptake, fluorescent tags for biosensing, and reactive functional groups for bioconjugation; however, with no prior information about their structure, most modifications abolish their binding function.<sup>47</sup> We therefore hypothesized that the high-content sequencing information afforded by our DNA-encoded oligomer library would be a useful approach to explore the sequence-activity relationship (SAR) landscape of these molecules. To this end, the thrombin aptamer is a useful proof-of-concept, given that it is one of the few aptamers with available crystal structures. As such, we were able to compare the frequency of different modifications tolerated in each position of the G-quadruplex scaffold to the known structural interactions with the thrombin target.

The DNA sequencing analysis indicates that the two Ts at position 2 and 6 are found at high frequency in the final screened library (more than 50% of all sequences, black in Figure S24), suggesting that modifying these positions may negatively impact function. Indeed, the crystal structure (PDB ID: 1hao) confirms the need for these two Ts to form polar interactions with the amino acid residues of exosite I of thrombin.<sup>48</sup> In contrast, the DNA sequencing results indicate that the scaffold tolerates all six modifications nearly equally in positions 1, 4, and 5 – all locations where the G-quadruplex does not directly interact with thrombin<sup>49-51</sup> (Figure S24). Functional binding assays of synthesized alenomers containing modifications at these positions displayed high affinity binding to thrombin, with  $K_{dS}$  in the nanomolar range (Table S6), further confirming the functional utility of the sequencing-based SAR analysis. Surprisingly, positions 3 and 7 also tolerated several diverse new modifications, despite known importance of these positions in forming hydrophobic interactions with thrombin.<sup>48</sup> Previous post-SELEX SAR analyses show that altering these positions reduces binding;<sup>49</sup> however, the majority of the tolerated modifications identified in our screen include hydrophobic moieties such as **Nap**, **Phe**, and **Trp** at position 3 and **Trp** and **C12** at position 7 (Figure S24). That these positions can be effectively replaced with diverse unnatural, hydrophobic nucleoside monomers with maintenance of binding opens the space of aptamer exploration to bespoke SDPs with potentially ameliorated function<sup>45</sup>.

Our non-nucleosidic library, coupled to the detailed analysis of the modification landscape described herein, provides important insight into the chemical nature of novel oligomer binding interactions. We hypothesize that this technology will enable the exploration of unique alenomer-target interactions without the need for crystal structures or time-consuming individual post-modification studies.

### Alenomers display enhanced binding

To determine the utility of the alenomer library to identify novel oligomers, the most popular candidates from the next generation sequencing, **AlenA** and **AlenB**, were synthesized without the branching unit, primers, or DNA barcodes for further analysis. Using multiple binding assays,<sup>52</sup> we confirmed that both novel alenomers bound to thrombin with high affinity ( $K_{dS}$  in the low nanomolar range: Figures 6, S13-18). **AlenB** displayed a 3-fold higher affinity to thrombin compared to the original thrombin aptamer (Tables S5-S7). The kinetics assessment of aptamer

binding using surface plasmon resonance (SPR) indicated that the increase in affinity of the highly-chemically modified **AlenB** was due to a 10 times longer off rate ( $k_d$ ), with only a modestly longer on rate ( $k_a$ ) (Figure 6). Slow off-rates are desirable for developing aptamers for specific biosensing applications.<sup>32</sup> Importantly, model alenomer constructs that we rationally designed without library selection displayed poor binding to thrombin (Table S6, Figure S18), highlighting that the DNA-encoded library developed in this work provides a much needed high-throughput screening platform to rapidly identify high affinity functional oligomers with non-nucleosidic modifications.

### **Alenomers display enhanced nuclease stability**

We observed that some of our alenomers had enhanced nuclease stability, as compared to **TBA**. **AlenA** showed a similar short half-life in serum (<0.3 h, Figure S19) as the original aptamer (which shows degradation within 108 seconds *in vivo*<sup>53</sup>), whereas **AlenB** resulted in more than 10-fold improved stability (half-life > 5 h). To investigate which non-nucleosidic modifications were responsible for this increased stability, several new constructs containing the individual modification were synthesized and tested, revealing that the presence of the 3'**Nap** modification protected the oligomer from degradation (half-life = 4.5 h). Most post-SELEX modifications are generally aimed at increasing binding affinity<sup>45, 48, 49, 54</sup> or improving nuclease resistance<sup>55-62</sup> – here, we achieved both with a single library screen.

## **Conclusions**

We developed a synthetic method to generate DNA-encoded, unnatural, sequence-defined oligomers by solid-phase synthesis. These libraries contain 300,000 members of G-quadruplex oligomers, each with up to seven unnatural modifications, enabling the high-throughput identification of novel thrombin-binding alenomers with improved binding affinity and serum stability. By correlating each non-nucleosidic monomer with its unique DNA code, the efficacy of site-specific modifications could be individually assessed. The use of non-nucleosidic modifications in aptamers is unprecedented, holding great promise for the discovery and optimisation of synthetic target-recognition SDPs.

Our solid-state method of DNA-encoded library synthetic represents a new strategy in forming DELs, which are often synthesized through enzymatic ligation<sup>17, 63</sup> Our synthetic route offers opportunities to broaden the chemical space explored by DEL, since it is carried out in non-aqueous solvents and under an inert atmosphere. While this work leveraged numerous monomers developed in previous work,<sup>21</sup> future libraries could make use of additional phosphoramidites from the plethora that are commercially available or readily synthesized. Importantly, solid-phase synthesis allows the use of other non-DNA compatible organic coupling reactions, in contrast to traditional DEL.

This study also highlights the potential for combinatorial strategies to identify target-binding sequences with unnatural sequence-defined oligomers. We leveraged next-generation sequencing to profile the sequence landscape of multiple stepwise library elutions, revealing structure-activity relationships of functional oligomers. We showed the potential of DEL of sequence-defined oligo(phosphodiester)s in the context of potential ligand discovery. However, due to the flexibility of this method the DNA-encoded oligomer library strategy can be extended to other applications, such as the discovery of ribozyme or DNAzyme catalysts. New monomers may also allow the exploration of new reactions that expand the scope and efficiency of DNA-inspired catalysts and binding units.

## **Acknowledgements**

We are grateful to members of the Sleiman and McKeague labs for useful discussions and for 2bind GmbH for the MST binding measurements. H. F. S. and M. M. are thankful to the National Science and Engineering Research Council of Canada (NSERC), the Canada Foundation for Innovation (CFI), the Fonds de Recherche Nature et Technologies (FRQNT), the Canada Institute for Health Research (CIHR), and Canada Research Chairs Program. H.F.S. is also thankful to the Canada Council for the Arts for a Killam Fellowship and is a Cottrell Scholar of the Research Corporation. F.J.R. and S. H. thank NSERC for a Banting Fellowship and CREATE Programmed Molecules for Therapeutics, Sensing and Diagnostics respectively. S. H. is thankful for Yvonne and Andrew Koenig Scholarship (The Jewish Community Foundation of Montreal). D.R., D.S., and S.A. are thankful to NSERC. D.R. is thankful for a CREATE training program in

Bionanomachines (CTPB) scholarship and S.A. is thankful for a Canada Graduate Scholarship – Master’s (CGS M).

### **Author contributions**

H.F.S. and D.R conceived the project with input from M.M. H.F.S., D.R., M.M. and S.H. designed experiments. D.R. developed and analyzed the DNA-encoded design and library. D.R., M.D., and V.T. synthesized and analyzed the monomers. S.H. and M.M. performed the screening and alenomer binding analysis. D.S. developed the DNA barcode amplification, stability experiments and analysis. S.A. analyzed library sequencing data and synthesized alenomers. D.R., S.H., S.A., F.J.R., H.S., M.M. co-wrote the paper with input from all authors.

### **Competing financial interests**

The authors declare the following competing financial interest(s): a US provisional patent application has been filed for the design of alenomers.

### **Data and materials availability**

All data supporting the findings of this study are available in the main text or the supplementary materials. All data are available from the corresponding author upon reasonable request.

## **References**

1. Deng, Z.; Shi, Q.; Tan, J.; Hu, J.; Liu, S., Sequence-Defined Synthetic Polymers for New-Generation Functional Biomaterials. *ACS Mater. Lett.* **2021**, 3 (9), 1339-1356.
2. Lutz, J. F., Defining the Field of Sequence-Controlled Polymers. *Macromol. Rapid Commun.* **2017**, 38 (24), 1700582.
3. Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J., A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991**, 354 (6348), 82-84.
4. Pavia, M. R.; Sawyer, T. K.; Moos, W. H., The generation of molecular diversity. *Bioorganic Med. Chem. Lett.* **1993**, 3 (3), 387-396.
5. Chen, C.; Randall, L. A. A.; Miller, R. B.; Jones, A. D.; Kurth, M. J., "Analogous" Organic Synthesis of Small-Compound Libraries: Validation of Combinatorial Chemistry in Small-Molecule Synthesis. *J. Am. Chem. Soc.* **1994**, 116 (6), 2661-2662.



6. DeWitt, S. H.; Kiely, J. S.; Stankovic, C. J.; Schroeder, M. C.; Cody, D. M.; Pavia, M. R., "Diversomers": an approach to nonpeptide, nonoligomeric chemical diversity. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90* (15), 6909-6913.
7. Cho, C. Y.; Youngquist, R. S.; Paikoff, S. J.; Beresini, M. H.; Hebert, A. R.; Berleau, L. T.; Liu, C. W.; Wemmer, D. E.; Keough, T.; Schultz, P. G., Synthesis and screening of linear and cyclic oligocarbamate libraries. Discovery of high affinity ligands for GPIIb/IIIa. *J. Am. Chem. Soc.* **1998**, *120* (31), 7706-7718.
8. Schneider, A. C.; Fritz, D.; Vasquez, J. K.; Vollrath, S. B.; Blackwell, H. E.; Bräse, S., Microwave-Facilitated SPOT-Synthesis of Antibacterial Dipeptoids. *ACS Comb. Sci.* **2017**, *19* (12), 715-737.
9. Celasun, S.; Remmler, D.; Schwaar, T.; Weller, M. G.; Du Prez, F.; Börner, H. G., Digging into the sequential space of thiolactone precision polymers: A combinatorial strategy to identify functional domains. *Angew. Chem. Int. Ed.* **2019**, *58* (7), 1960-1964.
10. Lokesh, G. L.; Wang, H.; Lam, C. H.; Thiviyanathan, V.; Ward, N.; Gorenstein, D. G.; Volk, D. E., X-Aptamer Selection and Validation. *Methods Mol Biol* **2017**, *1632*, 151-174.
11. Baldwin, J. J.; Burbaum, J. J.; Henderson, I.; Ohlmeyer, M. H. J., Synthesis of a Small Molecule Combinatorial Library Encoded with Molecular Tags. *J. Am. Chem. Soc.* **1995**, *117* (20), 5588-5589.
12. Martín, A.; Nicolaou, C. A.; Toledo, M. A., Navigating the DNA encoded libraries chemical space. *Commun. Chem.* **2020**, *3* (1), 127.
13. Brenner, S.; Lerner, R. A., Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89* (12), 5381-5383.
14. Boyle, J., Molecular Biology, 2nd Edition David P. Clark and Nanette J. Pazdernik, Academic Press-Cell, 2012, 928 pp., ISBN 978-0-1237-8594-7, \$135.00 (hardback). *Biochem. Mol. Biol. Educ.* **2014**, *42* (1), 100-101.
15. Clark, M. A., Selecting chemicals: the emerging utility of DNA-encoded libraries. *Curr. Opin. Chem. Biol.* **2010**, *14* (3), 396-403.
16. Goodnow, R. A.; Dumelin, C. E.; Keefe, A. D., DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov.* **2017**, *16* (2), 131-147.
17. Zhu, Z.; Shaginian, A.; La Shadric, C. G.; Davie, C. P.; Lind, K.; Pal, S.; Thansandote, P.; Simpson, G. L., DNA-encoded macrocyclic peptide library. In *Cyclic Peptide Design*, Springer: 2019; pp 273-284.
18. Usanov, D. L.; Chan, A. I.; Maianti, J. P.; Liu, D. R., Second-generation DNA-templated macrocycle libraries for the discovery of bioactive small molecules. *Nat. Chem.* **2018**, *10* (7), 704-714.
19. Beaucage, S.; Caruthers, M., Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **1981**, *22* (20), 1859-1862.
20. Shen, X.; Corey, D. R., Chemistry, mechanism and clinical status of antisense oligonucleotides and duplex RNAs. *Nucleic Acids Res.* **2018**, *46* (4), 1584-1600.
21. de Rochambeau, D.; Sun, Y.; Barlog, M.; Bazzi, H. S.; Sleiman, H. F., Modular Strategy To Expand the Chemical Diversity of DNA and Sequence-Controlled Polymers. *J. Org. Chem.* **2018**, *83* (17), 9774-9786.
22. Edwardson, T. G. W.; Carneiro, K. M. M.; Serpell, C. J.; Sleiman, H. F., An Efficient and Modular Route to Sequence-Defined Polymers Appended to DNA. *Angew. Chem. Int. Ed.* **2014**, *53* (18), 4567-4571.

23. de Rochambeau, D.; Barlóg, M.; Edwardson, T. G.; Fakhoury, J. J.; Stein, R. S.; Bazzi, H. S.; Sleiman, H. F., "DNA–Teflon" sequence-controlled polymers. *Polym. Chem.* **2016**, 7 (31), 4998-5003.
24. Al Ouahabi, A.; Kotera, M.; Charles, L.; Lutz, J.-F., Synthesis of Monodisperse Sequence-Coded Polymers with Chain Lengths above DP100. *ACS Macro Lett.* **2015**, 4 (10), 1077-1080.
25. Noir, R.; Kotera, M.; Pons, B.; Remy, J.-S.; Behr, J.-P., Oligonucleotide–Oligospermine Conjugates (Zip Nucleic Acids): A Convenient Means of Finely Tuning Hybridization Temperatures. *J. Am. Chem. Soc.* **2008**, 130 (40), 13500-13505.
26. Häner, R.; Garo, F.; Wenger, D.; Malinovskii, V. L., Oligopyrenotides: Abiotic, Polyanionic Oligomers with Nucleic Acid-like Structural Properties. *J. Am. Chem. Soc.* **2010**, 132 (21), 7466-7471.
27. Vasilyeva, S. V.; Budilkin, B. I.; Konevets, D. A.; Silnikov, V. N., Synthesis of Novel Nucleoside 5'-Triphosphates and Phosphoramidites Containing Alkyne or Amino Groups for the Postsynthetic Functionalization of Nucleic Acids. *Nucleosides Nucleotides Nucleic Acids* **2011**, 30 (10), 753-767.
28. Hébert, N.; Davis, P. W.; DeBaets, E. L.; Acevedo, O. L., Synthesis of N-substituted hydroxyprolinol phosphoramidites for the preparation of combinatorial libraries. *Tetrahedron Lett.* **1994**, 35 (51), 9509-9512.
29. Kupryushkin, M. S.; Nekrasov, M. D.; Stetsenko, D. A.; Pyshnyi, D. V., Efficient Functionalization of Oligonucleotides by New Achiral Nonnucleosidic Monomers. *Org. Lett.* **2014**, 16 (11), 2842-2845.
30. McLaughlin, C. K.; Hamblin, G. D.; Sleiman, H. F., Supramolecular DNA assembly. *Chem. Soc. Rev.* **2011**, 40 (12), 5647-5656.
31. Sefah, K.; Yang, Z.; Bradley, K. M.; Hoshika, S.; Jiménez, E.; Zhang, L.; Zhu, G.; Shanker, S.; Yu, F.; Turek, D., In vitro selection with artificial expanded genetic information systems. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, 111 (4), 1449-1454.
32. Gold, L.; Ayers, D.; Bertino, J.; Bock, C.; Bock, A.; Brody, E.; Carter, J.; Cunningham, V.; Dalby, A.; Eaton, B., Aptamer-based multiplexed proteomic technology for biomarker discovery. *Nat. Preced.* **2010**.
33. Hili, R.; Niu, J.; Liu, D. R., DNA ligase-mediated translation of DNA into densely functionalized nucleic acid polymers. *J. Am. Chem. Soc.* **2013**, 135 (1), 98-101.
34. Kong, D.; Yeung, W.; Hili, R., In vitro selection of diversely functionalized aptamers. *J. Am. Chem. Soc.* **2017**, 139 (40), 13977-13980.
35. Chen, Z.; Lichtor, P. A.; Berliner, A. P.; Chen, J. C.; Liu, D. R., Evolution of sequence-defined highly functionalized nucleic acid polymers. *Nat. Chem.* **2018**, 10 (4), 420-427.
36. Loakes, D.; Holliger, P., Polymerase engineering: towards the encoded synthesis of unnatural biopolymers. *Chem. Commun.* **2009**, (31), 4619-4631.
37. Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S. H., Synthetic genetic polymers capable of heredity and evolution. *Science* **2012**, 336 (6079), 341-344.
38. Pfeiffer, F.; Tolle, F.; Rosenthal, M.; Brändle, G. M.; Ewers, J.; Mayer, G., Identification and characterization of nucleobase-modified aptamers by click-SELEX. *Nat. Protoc.* **2018**, 13 (5), 1153.

39. Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H., Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature* **1991**, *354* (6348), 84-86.
40. Roxo, C.; Kotkowiak, W.; Pasternak, A., G-quadruplex-forming aptamers—characteristics, applications, and perspectives. *Molecules* **2019**, *24* (20), 3781.
41. Van Boom, J.; Burgers, P., Use of levulinic acid in the protection of oligonucleotides via the modified phosphotriester method: synthesis of decaribonucleotide uauauauaua. *Tetrahedron Lett.* **1976**, *17* (52), 4875-4878.
42. Horn, T.; Urdea, M. S., Forks and combs and DNA: the synthesis of branched oligodeoxyribonucleotides. *Nucleic Acids Res.* **1989**, *17* (17), 6959-6967.
43. Walton, T. A.; Lyttle, M. H.; Dick, D. J.; Cook, R. M., Evaluation of new linkers and synthetic methods for internal modified oligonucleotides. *Bioconjugate Chem.* **2002**, *13* (5), 1155-1158.
44. Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G., General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res.* **1991**, *37* (6), 487-493.
45. Dolot, R.; Lam, C. H.; Sierant, M.; Zhao, Q.; Liu, F.-W.; Nawrot, B.; Egli, M.; Yang, X., Crystal structures of thrombin in complex with chemically modified thrombin DNA aptamers reveal the origins of enhanced affinity. *Nucleic Acids Res.* **2018**, *46* (9), 4819-4830.
46. McKeague, M.; McConnell, E. M.; Cruz-Toledo, J.; Bernard, E. D.; Pach, A.; Mastronardi, E.; Zhang, X.; Beking, M.; Francis, T.; Giamberardino, A., Analysis of in vitro aptamer selection parameters. *J. Mol. Evol.* **2015**, *81* (5), 150-161.
47. Ni, S.; Zhuo, Z.; Pan, Y.; Yu, Y.; Li, F.; Liu, J.; Wang, L.; Wu, X.; Li, D.; Wan, Y., Recent progress in aptamer discoveries and modifications for therapeutic applications. *ACS Appl. Mater. Interfaces* **2020**, *13* (8), 9500-9519.
48. Tsvetkov, V. B.; Varizhuk, A. M.; Pozmogova, G. E.; Smirnov, I. P.; Kolganova, N. A.; Timofeev, E. N., A Universal Base in a Specific Role: Tuning up a Thrombin Aptamer with 5-Nitroindole. *Sci. Rep.* **2015**, *5* (1), 16337.
49. Platt, M.; Rowe, W.; Wedge, D. C.; Kell, D. B.; Knowles, J.; Day, P. J., Aptamer evolution for array-based diagnostics. *Anal. Biochem.* **2009**, *390* (2), 203-205.
50. Nallagatla, S. R.; Heuberger, B.; Haque, A.; Switzer, C., Combinatorial synthesis of thrombin-binding aptamers containing iso-guanine. *J. Comb. Chem.* **2009**, *11* (3), 364-369.
51. Smirnov, I.; Shafer, R. H., Effect of loop sequence and size on DNA aptamer stability. *Biochemistry* **2000**, *39* (6), 1462-1468.
52. McKeague, M.; De Girolamo, A.; Valenzano, S.; Pascale, M.; Ruscito, A.; Velu, R.; Frost, N. R.; Hill, K.; Smith, M.; McConnell, E. M., Comprehensive analytical comparison of strategies used for small molecule aptamer evaluation. *Anal. Chem.* **2015**, *87* (17), 8608-8612.
53. Griffin, L. C.; Tidmarsh, G. F.; Bock, L. C.; Toole, J. J.; Leung, L., In vivo anticoagulant properties of a novel nucleotide-based thrombin inhibitor and demonstration of regional anticoagulation in extracorporeal circuits. **1993**.
54. Ikebukuro, K.; Okumura, Y.; Sumikura, K.; Karube, I., A novel method of screening thrombin-inhibiting DNA aptamers using an evolution-mimicking algorithm. *Nucleic Acids Res.* **2005**, *33* (12), e108.
55. Sacca, B.; Lacroix, L.; Mergny, J.-L., The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.* **2005**, *33* (4), 1182-1192.

56. Martino, L.; Virno, A.; Randazzo, A.; Virgilio, A.; Esposito, V.; Giancola, C.; Bucci, M.; Cirino, G.; Mayol, L., A new modified thrombin binding aptamer containing a 5'-5' inversion of polarity site. *Nucleic Acids Res.* **2006**, *34* (22), 6653-6662.
57. Borbone, N.; Bucci, M.; Oliviero, G.; Morelli, E.; Amato, J.; D'Atri, V.; D'Errico, S.; Vellecco, V.; Cirino, G.; Piccialli, G., Investigating the role of T7 and T12 residues on the biological properties of thrombin-binding aptamer: enhancement of anticoagulant activity by a single nucleobase modification. *J. Med. Chem.* **2012**, *55* (23), 10716-10728.
58. Varizhuk, A. M.; Tsvetkov, V. B.; Tatarinova, O. N.; Kaluzhny, D. N.; Florentiev, V. L.; Timofeev, E. N.; Shchyolkina, A. K.; Borisova, O. F.; Smirnov, I. P.; Grokhovsky, S. L., Synthesis, characterization and in vitro activity of thrombin-binding DNA aptamers with triazole internucleotide linkages. *Eur. J. Med. Chem.* **2013**, *67*, 90-97.
59. Ying, G.; Lu, X.; Mei, J.; Zhang, Y.; Chen, J.; Wang, X.; Ou, Z.; Yi, Y., A structure-activity relationship of a thrombin-binding aptamer containing LNA in novel sites. *Bioorg. Med. Chem.* **2019**, *27* (14), 3201-3207.
60. Varada, M.; Aher, M.; Erande, N.; Kumar, V. A.; Fernandes, M., Methoxymethyl Threofuranosyl Thymidine (4'-MOM-TNA-T) at the T7 Position of the Thrombin-Binding Aptamer Boosts Anticoagulation Activity, Thermal Stability, and Nuclease Resistance. *ACS Omega* **2020**, *5* (1), 498-506.
61. Kovačič, M.; Podbevšek, P.; Tateishi-Karimata, H.; Takahashi, S.; Sugimoto, N.; Plavec, J., Thrombin binding aptamer G-quadruplex stabilized by pyrene-modified nucleotides. *Nucleic Acids Res.* **2020**, *48* (7), 3975-3986.
62. Riccardi, C.; Meyer, A.; Vasseur, J.-J.; Krauss, I. R.; Paduano, L.; Morvan, F.; Montesarchio, D., Fine-tuning the properties of the thrombin binding aptamer through cyclization: Effect of the 5'-3' connecting linker on the aptamer stability and anticoagulant activity. *Bioorg. Chem.* **2020**, *94*, 103379.
63. Favalli, N.; Bassi, G.; Scheuermann, J.; Neri, D., DNA-encoded chemical libraries—achievements and remaining challenges. *FEBS Lett.* **2018**, *592* (12), 2168-2180.