Enzymatic assembly of small synthetic genes with repetitive elements

Michael T. A. Nguyen[1,*], Martin Vincent Gobry[1], Néstor Sampedro Vallina[1], Georgios Pothoulakis[1], and Ebbe Sloth Andersen[1,2,*]

[1]Interdisciplinary Nanoscience Center (iNANO), Gustav Wieds Vej 14, Aarhus University, DK-8000 Aarhus, Denmark.

[2]Department of Molecular Biology and Genetics, Gustav Wieds Vej 14, Aarhus University, DK-8000 Aarhus, Denmark.

* To whom correspondence may be addressed:

Email: michaelnguyen91@live.dk, esa@inano.au.dk

Author Contributions: MN conceived the method, designed experiments, and analyzed the data. MN, MG, NS, and GP performed experiments. ESA and GP supervised. All authors discussed and wrote the paper.

Competing Interest Statement: No competing interests.

Keywords: Synthetic biology, DNA assembly methods, recombinant DNA technology, RNA nanotechnology, Molecular cloning.

This file includes:

Main Text

Figures 1-2

Table 1

Abstract

Gene synthesis efficiency has greatly improved in recent years but is limited when it comes to repetitive sequences and results in synthesis failure or delays by DNA synthesis vendors. This is representing a major obstacle for the development of synthetic biology, since repetitive elements are increasingly being used in design of genetic circuits and design of biomolecular nanostructures. Here, we describe a method for the assembly of small synthetic genes with repetitive elements: First, a gene of interest is split *in silico* into small synthons of up to 80 base pairs flanked by Golden Gate-compatible overhangs. Then, synthons are made by oligo extension and finally assembled into a synthetic gene by Golden Gate Assembly. We demonstrate the method by constructing eight challenging genes with repetitive elements e.g., multiple repeats of RNA aptamers and RNA origami scaffolds with multiple identical aptamers. The genes range in size from 133 to 456 base pairs and are assembled with fidelities of up to 87.5 %. The method was developed to facilitate our own specific research but may be of general use for constructing challenging and repetitive genes and thus a valuable addition to the molecular cloning toolbox.

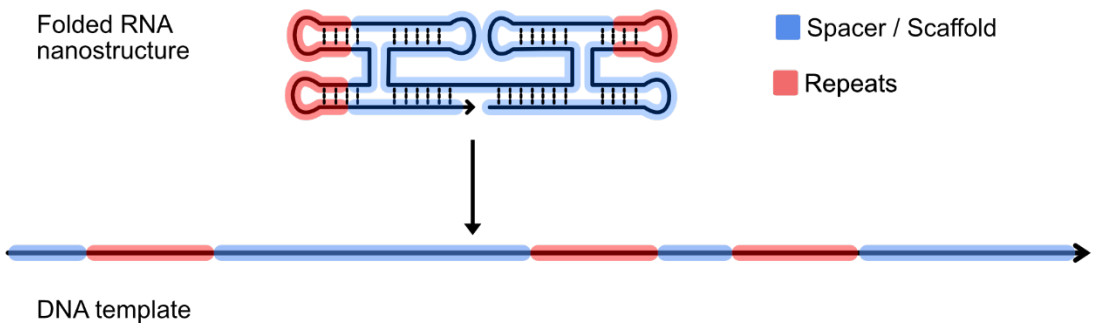
Introduction

Advances in DNA synthesis technologies have been pivotal for synthetic biology research. With the development of DNA assembly methods such as restriction-ligation[1], polymerase chain assembly[2], ligation-independent cloning methods[3, 4] and *in vitro*[5]- or *in vivo*[6]-homology-based methods, it is possible to construct almost any synthetic gene, biosynthetic pathway or genome. Of the many methods, the one-pot restriction-ligation method, Golden Gate Assembly[7] (GGA), has become a standard method for the synthetic biology workflows due to its utility for rapid and high-throughput cloning. GGA utilizes type

44 IIS restriction enzymes that cut distal to their recognition site, which allows overhangs to be designed
45 with arbitrary sequence for scarless cloning. Although DNA synthesis technologies have advanced
46 drastically, the synthesis of DNA sequences with high or low GC-content, stable secondary structures
47 and multiple repeats remain challenging[8], where especially repetitive DNA sequences result in
48 synthesis failure or delays by vendors[9]. To circumvent some of these challenges, it is possible to codon
49 optimize or codon scramble the sequence of genes encoding proteins to enable successful synthesis
50 with standard methods[10]. However, this approach is not applicable for genes encoding repeats of
51 functional non-coding RNA elements, since even a few sequence changes may change the folding and
52 negatively affect the function of the RNA element. Utilization of functional RNA elements, either isolated
53 from nature or *de novo* designed, has been increasingly popular in both the fields of synthetic biology[11]
54 and RNA nanotechnology[12]. Thus, a DNA assembly method to construct repetitive sequences is needed
55 for successful gene construction.

56 Methods to assemble synthetic DNA with repeats have been developed. These techniques assemble
57 shorter DNA sequences into larger repetitive DNA sequences using either restriction-ligation with type
58 IIS restriction enzymes in multiple steps[13, 14] or simply by annealing oligos[15] to create sequence-defined
59 overhangs for ligation. While these methods have proven useful for the scarless construction of
60 repetitive DNA, the approach with type IIS restriction enzymes can be put into a GGA workflow for rapid
61 assembly with minimal handling. GGA-based methods exist for the construction of at least 12 repeats
62 for CRISPR guide-RNA arrays[16]. However, these methods rely on pre-constructed templates for the
63 amplification of the scaffold part of the guide-RNA and are therefore not generalizable. Furthermore,
64 these methods have not demonstrated the assembly of repetitive DNA, where specific linker or scaffold
65 sequences with variable lengths are interspersed between the repeats such as for genes encoding RNA
66 origami nanostructures[12] that display multiple copies of the same RNA element (Figure 1). Therefore,
67 we wanted to build on this prior work and make a DNA assembly method to solve this challenge.

68



69

70 **Figure 1. Example of repetitive non-coding RNA gene.** Schematic representation of the secondary
71 structure of a 2-helix RNA origami scaffold[12] (blue) with repeated elements (red) and its associated
72 DNA template.

73

74 **Results and discussion**

75 Here, we describe a two-step DNA assembly method for the construction of synthetic genes with
76 repetitive elements that vendors could not provide. An overview of the method can be seen in Figure 2.
77 The method consists of splitting the gene of interest *in silico* into synthons of up to 80 base pairs (bp)
78 using Molecular Biology design software such as Benchling[17]. The genes are split into roughly equal-
79 sized fragments depending on the sequence size starting from position 1 to position 70-80 and then
80 from position 70-80 to position 140-150 and so on. We chose to design our method around GGA since
81 this method uses small overhangs (4 bp) that can be designed to minimize undesired annealing in
82 contrast to homology-based approaches which require at least 15-bp homology[4] and the synthons can
83 be directly inserted into a desired destination vector. We used the GGA tool from Benchling to design
84 compatible 4-bp overhangs. Other tools for overhang design exist such as NEB Golden Gate tool[18] or
85 the Golden Gate overhang designer from the Edinburgh biofoundry[19]. Synthons were made by oligo
86 extension with a DNA polymerase using oligos of up to 60 nucleotides (nt) that overlap with 18-25 bases
87 in their 3'-ends for hybridization to reduce misannealing and allow for extension of short oligos to longer
88 synthons. The synthetic gene is finally assembled and inserted into a desired entry vector by GGA.

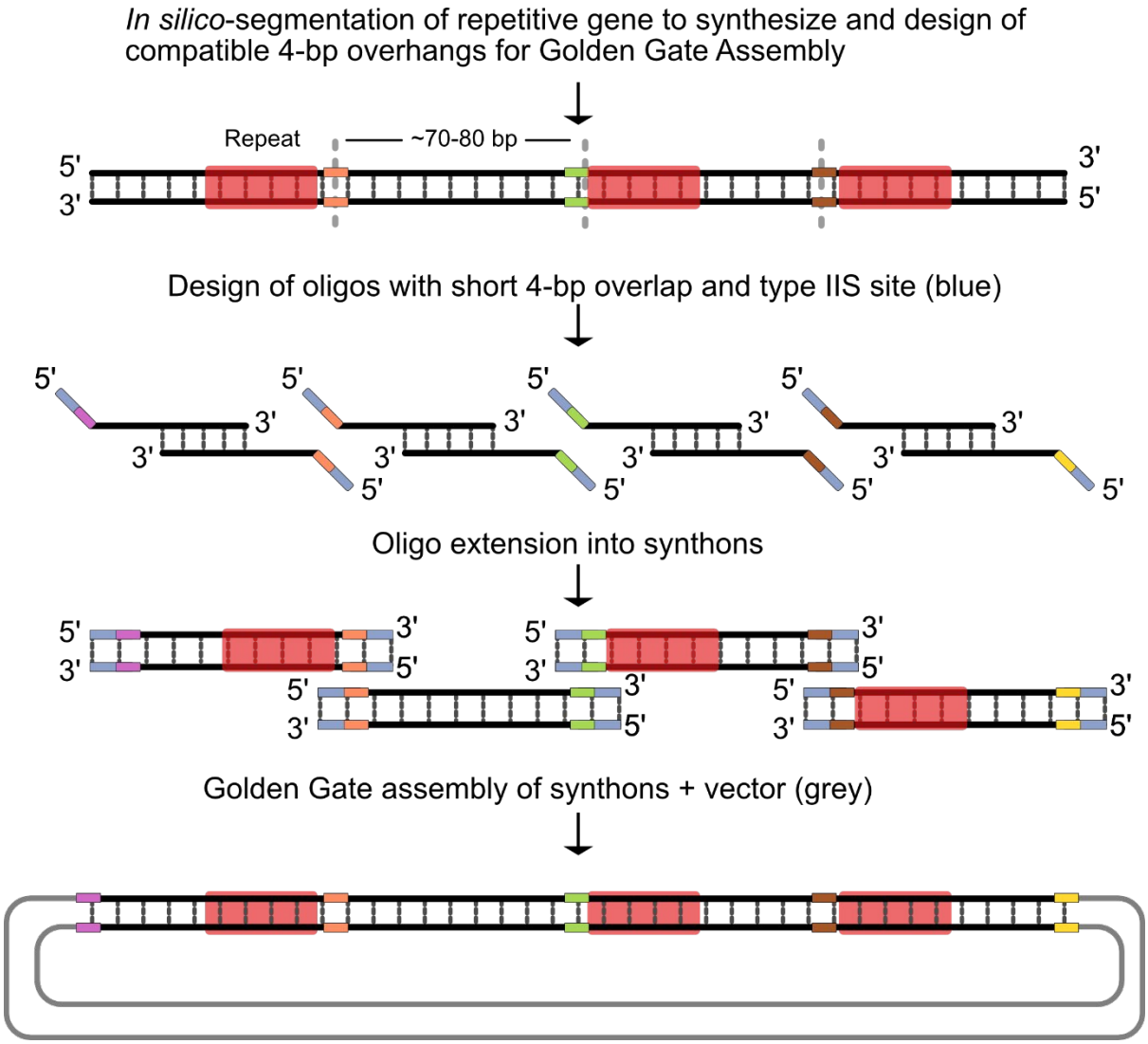


Figure 2. Overview of the two-step enzymatic synthesis of a synthetic gene. The designed gene with repeats (red) is split into 70-80 bp synthons *in silico* with 4-bp regions (orange, green, brown) that are used as overhangs for GGA. Oligos up to 60 nts are designed to cover the gene fragment and the oligos contain type IIS restriction enzyme sites in the 3'-ends (blue). Oligos covering the termini of the genes include overhangs (purple and yellow) compatible with an entry vector. Double-stranded DNA synthons are synthesized by oligo extension and inserted into a vector by GGA.

To demonstrate the applicability of our method, we assembled synthetic genes encoding tandem repeats of RNA aptamers with unique linker sequences between each repeat and RNA nanostructures with multiple repeated aptamers. These RNA constructs displayed different combinations of the known fluorogenic aptamers (Broccoli[20], Corn[21], Pepper[22]), the protein-binding aptamers (MS2[23] hairpin and Cas6f[24] hairpin) and a recently selected aptamer (P6Ba) that targets the SARS-CoV2 spike protein[25]. Details regarding their secondary structures (Figure S1) and sequences can be found in the supplementary material.

We chose these synthetic constructs because they require a well-defined sequence in between the repetitive elements and therefore cannot be synthesized by methods used for concatemerization of repeats[14]. Sequences ranging from 133 bp to 456 bp were designed. The sequences all failed initial screening for synthesis by the vendors Twist Bioscience and Integrated DNA Technologies[26] (IDT) due to their content of repeated elements with repeat lengths ranging from 20 bp to 81 bp. Furthermore, all the sequences have a complexity score from IDT ranging from 30 to 139, well above 10 (Table 1), which

is the maximum score allowed for synthesis. The sequence with the size of 133 bp was split into two fragments, the sequences ranging from 256 bp to 312 bp were split into four fragments, the sequence with 399 bp was split into five fragments and the two sequences with the sizes of 452 bp and 456 bp were split into six fragments. To synthesize the constructs, we used oligos of up to 60 nt, as this is the lowest price tier from IDT and each synthon was synthesized in individual reactions to reduce misannealing of the oligos that can cause faulty assembly. The synthons were assembled by GGA into different destination vectors. The constructs 2H-3xCorn and 2H-3xBroccoli were assembled with a linearized backbone, while the remaining constructs were assembled with a circular backbone. This was done due to the lack of availability of a desired circular entry backbone for those specific constructs. Plasmid sequences with accompanying sequencing data can be found in Table S2 and the sequences for the genes, synthons and oligos can be found in Table S3, Table S4, and Table S5, respectively.

We successfully constructed all the synthetic genes that could not be synthesized by the vendors (Table 1). For each construct, we sequenced at least eight clones with fidelity reported as the percentage of correct clones (Table 1). The highest observed fidelity of 87.5 % i.e., 7 correct clones out of 8 sequenced was observed for the assembly of 2x(Cas6f-hp-MS2wt) and 8xMS2.

Table 1. Summary of assembly fidelity for our eight different constructs with complexity scores from IDT.

Name	Length (bp)	Synthons	Complexity score	Fidelity (correct out of total)
2x(Cas6f-hp-MS2wt)	133	2	30.2	87.50% (7 out of 8)
2H-3xCorn	256	4	51.4	20.00% (2 out of 10)
8xCas6f-hp	280	4	139.2	62.50% (5 out of 8)
2H-3xBroccoli	301	4	103.3	62.50% (5 out of 8)
8xMS2	312	4	101	87.50% (7 out of 8)
PX-Tri-P6Ba-UU	399	5	88	25.00% (2 out of 8)
dAF14	452	6	123.9	12.50% (1 out of 8)
dAF10	456	6	112.9	12.50% (1 out of 8)

We observed that the synthesis fidelity decreased with increasing number of synthons. The 4-synthon constructs were assembled with a fidelity higher than 60 % except for one construct, 2H-3xCorn, where only 20 % of the selected clones were successful after two separate cloning attempts. For 2H-3xCorn, it appears that the linear backbone self-ligated, which resulted in colonies that only propagated the vector (Table S1). We observed that failed assemblies had a lack of partial or whole synthons (Table S1). The lack of partial synthons indicates errors in the oligo extension that results in shorter synthons while the lack of whole synthons indicates that the overhangs are suboptimal for GGA. This could potentially be improved with computational tools for data-optimized design of Golden Gate overhangs[27]. The failed assemblies could also be due to misannealing of the wrong synthons during the GGA reaction or due to manual mishandling of synthons when preparing the assembly in the lab e.g., using the wrong concentrations. We also observed failed assemblies with just single-bp mutations or deletions. These mutations or deletions could be either due to misannealing of oligos during the extension step, or due to synthesis errors. Prior work has shown that assembly of synthetic genes with chemically synthesized oligos can be improved after gel-purifying the oligos[28]. Furthermore, we did not sequence verify any synthons before assembling the final synthetic genes. Therefore, errors in the synthons could also have occurred during the oligo extension. However, by relying only on cheap and relatively short synthetic DNA oligos that have a short turnaround time, it is possible to synthesize genes faster compared to synthetic gene synthesis and delivery from commercial vendors.

Taken together, using simple design tools we demonstrate the construction of synthetic genes of up to 456 bp with repetitive elements that are not possible for common vendors to manufacture through their standard platform due to the sequence complexity. We report that the assembly fidelity dropped below 50 % with an increase in synthon number and complexity, low fidelity should not be a concern in cases where gene synthesis is otherwise impossible and efficiency levels should still be financially viable and non-laborious even at higher synthon numbers.

While this method was developed to facilitate our own specific research needs in the construction of genes encoding synthetic non-coding RNA devices and RNA nanostructures with repetitive elements, we envision it to not only be of use for RNA synthetic biology and RNA nanotechnology, but ultimately an important addition to the molecular cloning toolbox.

156 **Materials and Methods**

157 **RNA construct design**

158 2x(Cas6f-hp-MS2wt), 8xCas6f-hp, 8xMS2 were all designed using the NUPACK design package. The
159 secondary structures together with the sequences for the hairpins were used as design constraints.
160 Linker sequences were designed as non-repetitive sequences. The design with the lowest ensemble
161 score was chosen for synthesis. 2H-3xCorn, 2H-3xBr, PX-Tri-P6Ba-UU, dAF14, and dAF10 were all
162 designed using the ROAD[12] software and method for RNA origami design.

163 **Design of fragments and oligos**

164 Genes of interest are divided into 70-80 bp fragments *in silico* using Benchling. Compatible 4-bp
165 overhangs and initial primers with restriction site are designed with the Golden Gate assembly tool from
166 Benchling. Initial designed oligos are then manually extended in the Benchling sequence editor to
167 overlap by 18-25 bp for optimal annealing. Oligos are kept at a maximum length of 60 bases.

168 **Oligo extension for dsDNA synthesis**

169 DNA oligos (Integrated DNA technologies) for synthons were ordered resuspended in 1x IDT TE buffer.
170 Synthons were assembled by oligo extension using Q5 DNA Polymerase (NEB) in 25 µL reactions
171 consisting of 1x Q5 reaction buffer, 200 µM of each dNTP, 0.5 unit Q5 DNA polymerase, 200 nM oligos
172 (100 nM of each). Thermocycling was performed as following: initial denaturation at 98 °C for 30 s, then
173 six cycles of: denaturation at 98 °C for 10 s, annealing at synthon-dependent temperatures for 20 s,
174 elongation at 72 °C for 10 s, ending with a final elongation at 72 °C for 2 minutes. Synthons were purified
175 with the Nucleospin Gel and PCR clean-up kit (Macherey-Nagel) following the manufacturer's protocol
176 using a vacuum manifold. Synthon concentration was measured with a Denovix-11. Purified synthons
177 were diluted to 100 nM.

178

179 **Golden Gate reaction**

180 Golden Gate reactions were performed with equimolar amounts of DNA using 25 or 50 femtomoles of
181 synthon DNA, 25 femtomoles vector DNA, 0.5 µL T4 DNA ligase (NEB), 0.5 µL of either Esp3I (NEB)
182 or BsaI (NEB) in 1x T4 DNA ligase buffer with 10 µM ATP (NEB) in 10 µL reactions. Golden Gate
183 reactions consisted of 10 min at 37 °C, followed by 15 cycles of 5 min at 37 °C and 5 min at 16 °C
184 followed by heat-inactivation of the enzymes by a 5 minutes incubation at 50 °C and at 80 °C.

185 The whole reaction was transformed into NEB Turbo cells following standard protocols and cells were
186 plated on LB-agar plates containing either 100 µg/mL carbenicillin or 34 µg/ml chloramphenicol. Eight
187 to ten non-fluorescent colonies for each construct were picked for plasmid propagation. Plasmids were
188 propagated in cultures of terrific broth medium with appropriate antibiotics overnight at 37 °C and
189 purified with NucleoSpin miniprep kit (Macherey-Nagel) following the manufacturer's prototol. Plasmids
190 were verified by Sanger sequencing (Eurofins Genomics).

191

192 **Acknowledgments**

193 This project was financed by a Novo Nordisk Foundation Ascending Investigator grant (0060694). We
194 thank Cody Geary for designing the PX-Tri-6Ba construct for our test. Thanks to Rita Rosendahl and
195 Claus Bus for technical assistance.

References

(1) Cohen, S. N.; Chang, A. C. Y.; Boyer, H. W.; Helling, R. B. Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **1973**, *70* (11), 3240-3244. DOI: 10.1073/pnas.70.11.3240.

(2) Stemmer, W. P.; Crameri, A.; Ha, K. D.; Brennan, T. M.; Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **1995**, *164* (1), 49-53. DOI: 10.1016/0378-1119(95)00511-4 From NLM Medline.

(3) Quan, J.; Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* **2009**, *4* (7), e6441-e6441. DOI: 10.1371/journal.pone.0006441.

(4) Jeong, J. Y.; Yim, H. S.; Ryu, J. Y.; Lee, H. S.; Lee, J. H.; Seen, D. S.; Kang, S. G. One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. *Appl Environ Microbiol* **2012**, *78* (15), 5440-5443. DOI: 10.1128/AEM.00844-12 From NLM Medline.

(5) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A., 3rd; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **2009**, *6* (5), 343-345. DOI: 10.1038/nmeth.1318.

(6) Gibson, D. G.; Benders, G. A.; Axelrod, K. C.; Zaveri, J.; Algire, M. A.; Moodie, M.; Montague, M. G.; Venter, J. C.; Smith, H. O.; Hutchison, C. A., 3rd. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc Natl Acad Sci U S A* **2008**, *105* (51), 20404-20409. DOI: 10.1073/pnas.0811011106 From NLM Medline.

(7) Engler, C.; Kandzia, R.; Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS ONE* **2008**, *3* (11), e3647-e3647. DOI: 10.1371/journal.pone.0003647.

(8) Czar, M. J.; Anderson, J. C.; Bader, J. S.; Peccoud, J. Gene synthesis demystified. *Trends Biotechnol* **2009**, *27* (2), 63-72. DOI: 10.1016/j.tibtech.2008.10.007 From NLM Medline.

(9) Halper, S. M.; Hossain, A.; Salis, H. M. Synthesis Success Calculator: Predicting the Rapid Synthesis of DNA Fragments with Machine Learning. *ACS Synth Biol* **2020**, *9* (7), 1563-1571. DOI: 10.1021/acssynbio.9b00460 From NLM Medline.

(10) Tang, N. C.; Chilkoti, A. Combinatorial codon scrambling enables scalable gene synthesis and amplification of repetitive proteins. *Nat Mater* **2016**, *15* (4), 419-424. DOI: 10.1038/nmat4521 From NLM Medline.

(11) Chappell, J.; Takahashi, M. K.; Lucks, J. B. Creating small transcription activating RNAs. *Nat Chem Biol* **2015**, *11* (3), 214-220. DOI: 10.1038/nchembio.1737.

(12) Geary, C.; Grossi, G.; McRae, E. K. S.; Rothemund, P. W. K.; Andersen, E. S. RNA origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds. *Nat Chem* **2021**, *13* (6), 549-558. DOI: 10.1038/s41557-021-00679-1.

(13) Scior, A.; Preissler, S.; Koch, M.; Deuerling, E. Directed PCR-free engineering of highly repetitive DNA sequences. *BMC Biotechnol* **2011**, *11*, 87. DOI: 10.1186/1472-6750-11-87 From NLM Medline.

(14) McDaniel, J. R.; Mackay, J. A.; Quiroz, F. G.; Chilkoti, A. Recursive directional ligation by plasmid reconstruction allows rapid and seamless cloning of oligomeric genes. *Biomacromolecules* **2010**, *11* (4), 944-952. DOI: 10.1021/bm901387t From NLM Medline.

(15) Hamblin, G. D.; Rahbani, J. F.; Sleiman, H. F. Sequential growth of long DNA strands with user-defined patterns for nanostructures and scaffolds. *Nat Commun* **2015**, *6*, 7065. DOI: 10.1038/ncomms8065 From NLM Medline.

(16) McCarty, N. S.; Shaw, W. M.; Ellis, T.; Ledesma-Amaro, R. Rapid Assembly of gRNA Arrays via Modular Cloning in Yeast. *ACS Synth Biol* **2019**, *8* (4), 906-910. DOI: 10.1021/acssynbio.9b00041 From NLM Medline.

(17) *Benchling*. https://www.benchling.com/ (accessed.

(18) *NEBridge Ligase Fidelity*. https://ligasefidelity.neb.com/ (accessed.

(19) *Edinburgh Genome Foundry Overhang Designer*. https://cuba.genomefoundry.org/design-overhangs/ (accessed.

(20) Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R. Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society* **2014**, *136* (46), 16299-16308. DOI: 10.1021/ja508478x.
(21) Song, W.; Filonov, G. S.; Kim, H.; Hirsch, M.; Li, X.; Moon, J. D.; Jaffrey, S. R. Imaging RNA polymerase III transcription using a photostable RNA-fluorophore complex. *Nat Chem Biol* **2017**, *13* (11), 1187-1194. DOI: 10.1038/nchembio.2477 From NLM Medline.
(22) Chen, X.; Zhang, D.; Su, N.; Bao, B.; Xie, X.; Zuo, F.; Yang, L.; Wang, H.; Jiang, L.; Lin, Q.; et al. Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nature Biotechnology* **2019**. DOI: 10.1038/s41587-019-0249-1.
(23) Peabody, D. S. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J* **1993**, *12* (2), 595-600.
(24) Haurwitz, R. E.; Jinek, M.; Wiedenheft, B.; Zhou, K.; Doudna, J. A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **2010**. DOI: 10.1126/science.1192272.
(25) Valero, J.; Civit, L.; Dupont, D. M.; Selnihhin, D.; Reinert, L. S.; Idorn, M.; Israels, B. A.; Bednarz, A. M.; Bus, C.; Asbach, B.; et al. A serum-stable RNA aptamer specific for SARS-CoV-2 neutralizes viral entry. *Proc Natl Acad Sci U S A* **2021**, *118* (50). DOI: 10.1073/pnas.2112942118 From NLM Medline.
(26) *Integrated DNA Technologies*. https://www.idtdna.com/ (accessed.
(27) Pryor, J. M.; Potapov, V.; Bilotti, K.; Pokhrel, N.; Lohman, G. J. S. Rapid 40 kb Genome Construction from 52 Parts through Data-optimized Assembly Design. *ACS Synth Biol* **2022**, *11* (6), 2036-2042. DOI: 10.1021/acssynbio.1c00525 From NLM Medline.
(28) Smith, H. O.; Hutchison, C. A., 3rd; Pfannkoch, C.; Venter, J. C. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci U S A* **2003**, *100* (26), 15440-15445. DOI: 10.1073/pnas.2237126100.