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

Figure 1

Table 1

Abstract

Although gene synthesis efficiency has improved through the years, current methods offered by DNA synthesis vendors are limited when it comes to repetitive sequences. Here, we describe a method for the enzymatic assembly of repetitive small synthetic genes. The method involves an initial step where the gene of interest is split *in silico* into small synthons of up to 80 base pairs flanked by Golden Gate-compatible four-base pair overhangs. Synthons are enzymatically synthesized by oligo extension and then assembled into the synthetic gene by Golden Gate Assembly. We construct eight different synthetic genes ranging from 133 to 456 base pairs encoding RNA structures with repetitive elements that are challenging to synthesize. We report assembly fidelities of up to 87.5 % that decrease for increasing number of synthons. This method is envisioned as an important addition to the molecular cloning toolbox and especially useful for construction of challenging and repetitive genes.

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 IIS restriction enzymes that cut distal to their recognition site, which allows overhangs to be designed with arbitrary sequence for scarless cloning. Although DNA synthesis technologies have advanced drastically, the synthesis of DNA sequences with high or low GC-content, stable secondary structures and multiple repeats remain challenging[8], where especially

repetitive DNA sequences result in synthesis failure or delays by vendors[9]. To circumvent some of these challenges it is possible to codon optimize or codon scramble the sequence of genes encoding proteins to enable successful synthesis with standard methods[10]. However, this approach is not applicable for genes encoding repeats of functional non-coding RNA elements, since even a few sequence changes may change the folding and negatively affect the function of the RNA element. Thus, a DNA assembly method to construct repetitive sequences is needed for successful gene construction. A few methods to assemble synthetic DNA with repeats have been developed. These techniques assemble shorter DNA sequences into larger repetitive DNA sequences using either restriction-ligation with type IIS restriction enzymes in multiple steps[11, 12] or simply by annealing oligos[13] to create sequence-defined overhangs for ligation. While these methods have proven useful for the scarless construction of repetitive DNA, the approach with type IIS restriction enzymes can be put into a GGA workflow for rapid assembly with minimal handling. Furthermore, these methods have not demonstrated the assembly of repetitive DNA, where specific linker or scaffold sequences are interspersed between the repeats such as for genes encoding RNA nanostructures that display multiple copies of the same RNA element. Therefore, we wanted to build on this prior work and make a DNA assembly method to circumvent this challenge.

Results and discussion

Here, we describe a two-step DNA assembly method for the construction of synthetic genes with repetitive elements that vendors could not provide. An overview of the method can be seen in Figure 1. The method consists of manually splitting the gene of interest *in silico* into synthons of up to 80 base pairs (bp) using Molecular Biology design software such as Benchling[14]. The genes are split into roughly equal-sized fragments depending on the sequence size starting from position 1 to position 70-80 and then from position 70-80 to position 140-150 and so on. We used the GGA tool from Benchling to design compatible 4-bp overhangs. Other tools for overhang design exist such as NEB Golden Gate tool[15] or the Golden Gate overhang designer from the Edinburgh biofoundry[16]. We chose to design our method around GGA since this method uses small overhangs (4 bp) that can be designed to minimize undesired annealing in contrast to homology-based approaches which require at least 15-bp homology[4] and the synthons can be directly inserted into a desired destination vector. Synthons were made by oligo extension with a DNA polymerase using oligos of up to 60 nucleotides (nt) that overlap with 18-25 bases in their 3'-ends for hybridization to reduce misannealing and allow for extension of short oligos to longer synthons. The synthetic gene is finally assembled and inserted into a desired entry vector using 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[17], Corn[18], Pepper[19], the protein-binding aptamers MS2[20] hairpin and Cas6f[21] hairpin and a recently selected aptamer (P6Ba) that targets the SARS-CoV2 spike protein[22]. 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[12]. Sequences ranging from 133 bp to 456 bp were designed. The sequences all failed initial screening for synthesis by the vendors Twist Bioscience and 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 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 Integrated DNA Technologies[23] (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. 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 synthesized all constructs that could not be synthesized by the vendors (Table 1). For each construct, we sequenced at least eight clones and the fidelity is reported as the percentage of correct clones (Table 1). The highest observed fidelity was 87.5 % i.e. 7 correct clones out of 8 sequenced, which was observed for the assembly of 2x(Cas6f-hp-MS2wt) and 8xMS2.

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. 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 impurities of the oligos. Prior work has shown that assembly of synthetic genes with chemically synthesized oligos can be improved after gel-purifying the oligos[24].

Our design approach for the assemblies was done using simple design tools, which seem to at least be applicable for constructs with up to six synthons. However, with the increase in synthon number and complexity, we observed that the assembly fidelity dropped to below 50 %, which could be due to lower GGA efficiency that can be improved with the recent computational tools for data-optimized design of Golden Gate overhangs[25]. We also did not sequence verify any synthons before assembling the final synthetic genes. Therefore, errors in the synthons could have occurred during the oligo extension.

Taken together, we managed to assemble synthetic genes that are not possible for common vendors to manufacture through their standard platforms. Furthermore, by relying on cheap and relatively short synthetic DNA oligos that have a short turnaround time, this method can possibly be used to synthesize genes faster compared to synthetic gene synthesis and delivery from commercial vendors. In conclusion, we demonstrate the possibility to synthesize repetitive genes up to 456 bp that are challenging to synthesize. This work can be applied to construct synthetic CRISPR spacer arrays with multiple repeats, RNA nanostructures that incorporate multiples of the same aptamer or synthetic genes encoding repeated protein sequences that cannot be synthesized even after codon optimization. Ultimately, we envision this work to benefit RNA synthetic biology and RNA nanotechnology and that the method can become an important addition to the molecular cloning toolbox.

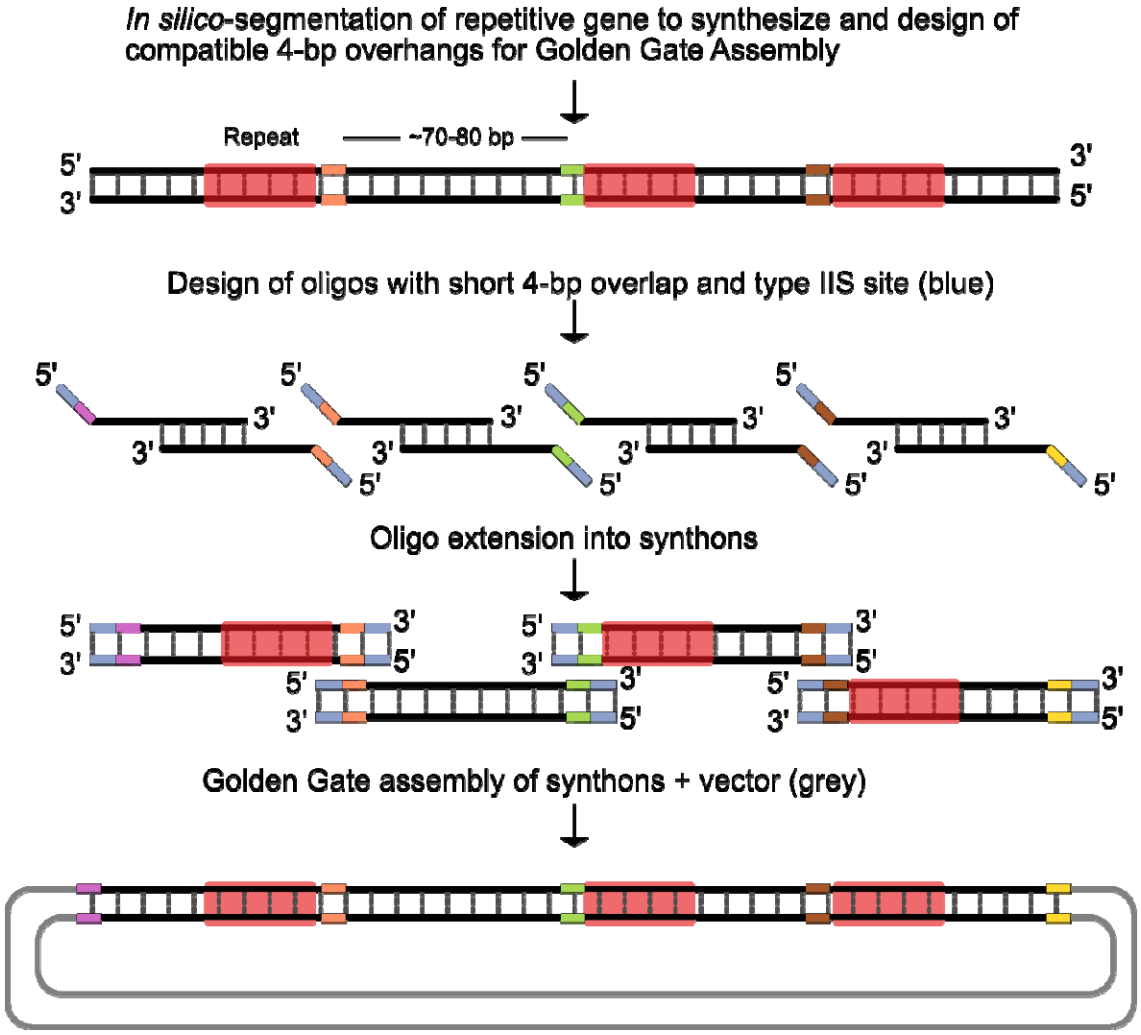

Figure 1. Overview of the two-step enzymatic synthesis of a synthetic gene. The designed gene with repeats (red) is split into 70-80 bp synthons *in silico* with 4-bp regions (orange, green, brown) that are used as overhangs for GGA. Oligos up to 60 nts are designed to cover the gene fragment and the 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.

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)

Materials and Methods

RNA construct design

2x(Cas6f-hp-MS2wt), 8xCas6f-hp, 8xMS2 were all designed using the NUPACK design package. The secondary structures together with the sequences for the hairpins were used as design constraints. Linker sequences were designed as non-repetitive sequences. The design with the lowest ensemble score was chosen for synthesis. 2H-3xCorn, 2H-3xBr, PX-Tri-P6Ba-UU, dAF14, and dAF10 were all designed using the ROAD[26] software and method for RNA origami design.

Design of fragments and oligos

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

Oligo extension for dsDNA synthesis

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

Golden Gate reaction

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

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

Acknowledgments

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

(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.

(3) Quan, J.; Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* **2009**, *4* (7), e6441-e6441.

(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.

(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.

(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.

(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.

(8) Czar, M. J.; Anderson, J. C.; Bader, J. S.; Peccoud, J. Gene synthesis demystified. *Trends Biotechnol* **2009**, *27* (2), 63-72.

(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.

(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.

(11) Scior, A.; Preissler, S.; Koch, M.; Deuerling, E. Directed PCR-free engineering of highly repetitive DNA sequences. *BMC Biotechnol* **2011**, *11*, 87.

(12) 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.

(13) 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.

(14) *Benchling*. https://www.benchling.com/.

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

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

(17) 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.

(18) 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.

(19) 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**.

(20) Peabody, D. S. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J* **1993**, *12* (2), 595-600.

(21) Haurwitz, R. E.; Jinek, M.; Wiedenheft, B.; Zhou, K.; Doudna, J. A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **2010**.

(22) 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).

(23) *Integrated DNA Technologies*. https://www.idtdna.com/.

(24) 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.

(25) 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.

(26) 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.