



Supporting Information

© Wiley-VCH 2012

69451 Weinheim, Germany

RNA-Mediated Gene Assembly from DNA Arrays**

*Cheng-Hsien Wu, Matthew R. Lockett, and Lloyd M. Smith**

anie_201109058_sm_miscellaneous_information.pdf

Supporting Information

Materials and Methods

Design of DNA arrays for RNA-mediated assembly

The full-length ZsGreen1 coding gene plus Kozak sequence (696 nt RNA transcript encoding ZsGreen1, and 10 additional nucleotides corresponding to the Kozak sequence for eukaryotic cell-free expression system) was split into 18 segments. Segments were designed to include terminal GG dinucleotides to enhance *in vitro* T7 RNA polymerase transcription. Every segment is longer than 30 nt to provide a minimum of 15 bp hybridization with the splint oligos. Splint RNAs were appended with an initial GG dinucleotide for the same reason as for the target fragments.

37 single stranded DNAs were synthesized on the microarray (see sequence information for details), which includes 18 segment RNA templates, 17 splint RNA templates, and two control oligos for quality monitoring. For each of the 9 longest segment sequences (>67mer), multiple features were made rather than just one, in order to increase the amount of RNA produced (see “sequences on the microarray” below). Each feature is sized 1680 μm x 1232 μm .

Preparation of substrates for *in situ* photolithographic oligonucleotide array synthesis

Silanized glass. Glass is the standard substrate for DNA array fabrication because of its advantages of low intrinsic fluorescence, non-porosity and ease of modification using silane chemistries. Glass microscope slides (Plain Micro Slides, VWR, PA, USA) were cleaned with 1M sodium hydroxide prior to silanization. The slides were then silanized for 4h in 2% (v/v) N-(2-triethoxysilylpropyl)-4-hydroxybutyramide (Gelest, Inc., Morrisville, PA, USA) in stock solution (0.1% acetic acid in 95% ethanol). After being rinsed by stirring in fresh stock solution for 15 min, the slides were transferred to a pre-heated (120°C) oven for 2h, and cured under vacuum overnight.

Carbon-on-gold. In addition to the use of the above standard glass substrates for DNA array fabrication, we also employed substrates overlaid with amorphous carbon deposited on a gold thin film because of their superior thermal stability [8b]. Tetraethylene glycol monoallylether was employed for the preparation of hydroxyl terminated surfaces for photolithographic oligonucleotide array synthesis since it has been reported that polyethylene glycol modified surfaces help to reduce nonspecific adsorption of proteins [18]. First, standard glass slides coated with 50Å chromium and 1,000 Å of gold (EMF corp., NY, USA) were extensively rinsed with hexane and ethanol and dried under a nitrogen stream. A 7.5 nm layer of amorphous carbon was then DC magnetron sputtered on the gold surface (Denton Vacuum, NJ, USA). 40 μL of tetraethyleneglycol monoallylether, which was synthesized according to a literature procedure [19], was placed directly onto the amorphous carbon surface, and then covered with a quartz coverslip. The surfaces were irradiated under nitrogen purge with a low-pressure mercury vapor quartz grid lamp ($\lambda = 254 \text{ nm}$, 0.35 mW/cm²) for 16 h. After the photoreaction, the surfaces were rinsed extensively with ethanol and deionized water and dried under nitrogen.

In situ photolithographic oligonucleotide array synthesis

Light-directed photolithographic synthesis of DNA arrays was performed on either the silanized glass slides or the ethylene glycol modified carbon-on-gold surface with a digital micromirror-based Maskless Array Synthesis (MAS) system connected to a ABI Expedite™ 8909 Nucleic Acid Synthesis System (Applied Biosystems, CA, USA) as described previously [8a,11a,20]. Oligonucleotide synthesis reagent, 0.1M Activator 42 (5-[3,5-Bis(trifluoromethyl)phenyl]-1H-tetrazole) and all NPPOC (3'- nitrophenylpropyloxycarbonyl) protected phosphoramidites [5'- NPPOC-dAdenosine (tac) 3'- β -cyanoethylphosphorimidite (NPPOC-dA), 5'- NPPOC-dThymidine 3'- β -cyanoethylphosphorimidite (NPPOC-dT), 5'- NPPOC-dCytidine (ib) 3'- β -cyanoethylphosphorimidite (NPPOC-dC), 5'- NPPOC-dGuanosine (ipac) 3'- β -cyanoethylphosphorimidite (NPPOC-dG)] were purchased from Sigma Aldrich. Anhydrous wash (acetonitrile), amidite diluent (acetonitrile), capping reagent A (THF/PAC₂O), capping reagent B (Cap Mix B 10% N-Methylimidazole in THF) and deblocking mix (3% dichloroacetic acid in dichloromethane) were purchased from Glen Research (VA, USA). Oxidizing reagent (0.02 M iodine in THF/pyridine/H₂O, 89.6/0.4/10) was purchased from EMD Chemicals (NJ, USA). Exposure solvent is 1% imidazole in DMSO. Anhydrous reagents were kept over molecular sieves (AldraSORB™ water trapping packets, Sigma Aldrich). The oligonucleotide synthesis protocol was modified and optimized based on previously published protocols. [8b,9b] Briefly, every synthesis cycle contains two capping steps to achieve high yield of full-length templates and one oxidation to stabilize the phosphodiester bonds. The step-sequence was coupling (80 sec), capping (20 sec), oxidizing (15 sec), capping (flow though), and UV deprotection. The light dose to remove the photolabile NPPOC

(3'-nitrophenylpropyloxycarbonyl)-protecting groups from NPPOC phosphoramidites (Sigma Aldrich, MO, USA) was determined prior to DNA array fabrication. A series of incremental doses of 365 nm light (Joule/cm^2) was used for a 30 nt quality control (QC) oligonucleotide synthesis. The optimal dose was chosen to yield the highest level of fluorescence from hybridization of a fluorescently tagged QC complement. It is noted that the complete removal of NPPOC protecting group is important to eliminate possible deletions during synthesis. Array synthesis proceeded as follows: (a) after coupling of the previous NPPOC-protected base to the growing DNA strand, the synthesis flow cell (volume~100 μl) was flushed with 500 μl of exposure solvent; (b) a digital image (mask) representing the locations for the next base addition illuminated the surface with either 4.2 Joule/ cm^2 of 365 nm light on silanized glass or 3.5 Joule/ cm^2 of 365 nm on carbon surface using a 350 watt mercury arc lamp (Newport, CT, USA). Exposure solvent was constantly flowed through the flow cell at a rate of 180 ~ 220 $\mu\text{l}/(\text{Joule}/\text{cm}^2)$ during illumination, sufficiently maintaining the basic conditions needed to drive the photocatalyzed elimination reaction.^[21] (c) Following irradiation, the array was washed with acetonitrile (~400 μl) to remove residual exposure solvent, dry wash (~300 μl) to remove trace water, and activator solution (~100 μl). (d) Coupling of the next base was achieved by filling the flow cell with a 1:1 solution of the desired phosphoramidite and Activator 42. All 5'-NPPOC-protected amidites underwent a single 80s coupling step. (e) After amidite coupling, the array was capped with a 1:1 v/v mixture of capping reagents A and B (A:B solution) for 20 sec (~320 μl). (d) After washing with acetonitrile (~100 μl) the array was oxidized with oxidizer solution for 15 sec (THF, pyridine, iodine, and water; ~480 μl). (e) A second capping step was performed by flushing the cell with capping reagent A:B solution. (f) After synthesis is complete, the nucleoside bases are deprotected in 1:1 ethylenediamine:absolute ethanol solution at room temperature for 2 hr. The primary significant differences from previously published protocols^[8b,9b] are: (i) a higher photo dose was used to remove the NPPOC-protecting groups on the carbon-on-gold surface; (ii) a longer coupling time (80 sec) and different activator (Activator 42) were used; (iii) capping was conducted directly after each amidite coupling followed by oxidation and another capping step; (iv) An oxidizing step was included in every cycle.

On-chip RNA transcription with T7 RNA polymerase

A gasket, Gene Frame – 15 x 16mm internal (Abgene, Epsom, UK), was attached and surrounded the DNA features. A 100 μl total *in vitro* RNA transcription reaction contains a final concentration of 2.25 U/ μl T7 RNA polymerase-plusTM (Ambion, TX, USA), 0.8 mM each NTP, 1 μM T7RNAP promoter complement, 1x RNAsecureTM reagent (Ambion, TX, USA), 20mM NaCl, 40 mM Tris pH 7.8, 6 mM MgCl₂, 2 mM spermidine, and 10 mM DTT. The reaction mixture, except T7 RNA polymerase, was applied to the chip and incubated at 60°C for 10 min, then slowly cooled down to room temperature. T7 RNA polymerase was then added to the surface. The transcription reaction was conducted at room temperature for 4~12 hr in a humid chamber. The total reaction was collected and desaltsed three times with deionized water using a cellulose-based 3,000 molecular weight cut-off Amicon Ultra-0.5mL centrifugal filter (Millipore, MA, USA) prior to pyrophosphate removal.

Pyrophosphate removal from triphosphorylated RNA transcripts

RNA transcripts initiated with triphosphorylated GG dinucleotides were treated with RNA 5' Pyrophosphohydrolase (RppH) (NEB, MA, USA) in amended T4 RNA ligase 2 reaction buffer (without ATP) instead of 1 x NEBuffer 2 (NEB) to reduce the possibility of losses due to extra steps, and to simplify the overall assembly process. 5 units of RppH were used to remove pyrophosphate group in a half volume of buffer exchanged RNA transcription reaction in a final concentration of 50 mM Tris-HCl pH7.5, 2mM MgCl₂, and 1 mM DTT. The reaction was incubated at 37°C for 2hr in a total volume of 25 μl .

Full-length RNA ligation with T4 RNA ligase 2

10 units of T4 RNA ligase 2 and a final concentration of 800 μM ATP were added to the RppH treated reaction above (a half of the total on chip transcribed RNAs.) The ligation reaction involved an initial ligation step at 37°C for 10 min, followed by 3 cycles of thermal-cycled ligation at 65°C for 30 sec and 37°C for 5 min, and finished with a final ligation step at 37°C for 10 min.

Reverse transcription PCR for assembled RNA transcripts

Assembled ZsGreen1 RNA transcripts for cloning and prokaryotic cell-free protein expression were amplified by reverse transcription PCR (RT-PCR) using a OneStep RT-PCR Kit (QIAGEN, CA, USA). ZsGreen1 specific primers, ZsG-F and ZsG-R-w-6His were used (see Sequence Information for details). Cycling consisted of 30 min at 50°C, 15 min at 95°C; then 40 cycles of 30 sec at 95°C, 30 sec at 61°C, and 1 min at 72°C; and final elongation 10 min at 72°C.

Assembled ZsGreen1 RNA transcripts for eukaryotic cell-free protein expression were amplified by RT-PCR using a GeneAmp Gold RNA PCR Reagent Kit (Applied Biosystems, CA, USA). ZsGreen1 specific primers: ZsG-F and ZsG-R were used. Cycling consisted of 12 min at 42°C, 10 min at 95°C; then 45 cycles of 20 sec at 94°C, 20 sec at 58°C, and 30 min at 72°C; and final elongation 7 min at 72°C. The ZsGreen1 DNA was gel purified. Next, T7-ZsG-F and ZsG-R primers were used to append a T7 promoter to ZsGreen1 coding gene. Phusion Hot Start High-Fidelity DNA Polymerase (NEB) was used. Cycling consisted of 30 sec at 98°C; then 35 cycles of 10 sec at 98°C, 20 sec at 62°C, and 30 sec at 72°C; and final elongation 10 min at 72°C.

The RT-PCR products were analyzed by electrophoresis in a 1.5% agarose gel along with a 100 bp DNA ladder (NEB).

Cell-free protein expression, purification and detection

1. Prokaryotic cell-free protein expression.

The assembled ZsGreen1 gene without Kozak sequence was ligated to pEXP5-CT/TOPO vector (Invitrogen, OR, USA) followed by transformation into ONE Shot TOP10 Competent *E. coli* cells (Invitrogen). The plasmids with inserts were purified with a QIAprep Spin Miniprep Kit (QIAGEN). One microgram of plasmid DNA was used in a standard 100 µl reaction of Expressway Mini Cell-Free *E. coli* Expression System (Invitrogen). The protein expression reaction was performed for 4 hr at 30°C. ZsGreen1 protein was either directly analyzed in protein gels or purified with Ni-NTA Magnetic Agarose Beads (QIAGEN) prior to the analysis.

2. Eukaryotic cell-free protein expression.

ZsGreen1 RNA transcripts with Kozak sequence were produced from assembled T7 promoter appended ZsGreen1 gene by using a MEGAscript T7 kit (Ambion). The transcription reactions were buffer exchanged with water using a cellulose-based 30,000 molecular weight cut-off Amicon Ultra-0.5mL centrifugal filter. Approximately 3.7 micrograms of RNA transcripts were used in a 20 µl Retic Lysate IVT (Ambion) cell-free expression reaction.

3. Protein analysis.

The protein products obtained from the *in vitro* expression system were analyzed in either reducing (a final concentration of 2.5% ? - mercaptoethanol was added to denature the samples at 95°C for 5 min), or non-reducing gradient SDS-PAGE gels (4-20%, Bio-Rad, Richmond, CA, USA). The prestained broad range protein standard marker (7-175 kDa) run along with the protein samples in the SDS-PAGE gel was purchased from NEB. The reducing SDS-PAGE gels were visualized by Coomassie Blue staining. The fluorescent proteins in the non-reducing SDS-PAGE gels were visualized under a 488 nm laser with a 530 nm filter using a Bio-Rad Molecular Imager FX Pro.

Sequence Information

Target sequence (ZsGreen1, adapted from Clontech's pZsGreen1-C1 vector)

Note: The underscored region is the Kozak sequence. The initial GG is included in the T7 RNAP transcript for better transcription efficiency.

```
GGTCGCCACCATGGCCCAGTCCAAGCACGGCTGACCAAGGAGATGACCATGAAGTACCGCATGGAGGGCTGCCTGGACGG
CCACAAGTTCTGTGATCACCAGCGAGGGCATCGGCTACCCCTCAAGGGCAAGCAGGCCATCACCTGTGCGTGGAGGGC
GGCCCCCTGCCCTCGCCGAGGACATCTGTCCGCCCTCATGTACGGCAACCGCGTGTTCACCGACTACCCCCAGGACAT
CGTCGACTACTTCAAGAACCTCTGCCCGCCGCTACACCTGGGACCGCTCCTCTGTTCGAGGACGGCGCCGTGTCATCT
GCAACGCCGACATCACCCTGAGCGTGAGGAGAACTGCATGTACACAGTCCAAGTTCTACGGCGTGAACTTCCCAGCGA
CGGCCCCGTGATGAAGAAGATGACCGACAACGGGAGCCCTCTGCGAGAAGATCATCCCCGTGCCAAGCAGGGCATCTT
AAGGGCGACGTGAGCATGTACCTGCTGCTGAAGGACGGTGCCGCTGCGCTGCCAGTTGACACCGTGTACAAGGCAAGT
CCGTGCCCGCAAGATGCCGACTGGCACTTCATCCAGCAGCTGACCCCGAGGACCGCAGCGACGCCAAGAACCAAGAA
GTGGCACCTGACCGAGCACGCCATGCCCTCGGCCCTGCCCTGA
```

Target segment-RNAs for ZsGreen1 assembly (5' to 3')

- 1 GGTCGCCACCATGGCCCAGTCCAAGCACGGCTGACCAA
- 2 GGAGATGACCATGAAGTACCGCATGGAGGGCTGCGT
- 3 GGACGGCCACAAGTTCTGTGATCACCGCGA
- 4 GGGCATCGGCTACCCCTCAAGGGCAAGCA
- 5 GGCCATCAACCTGTGCGTGGAGGGCGGCCCTTGCCCTCGCCGA
- 6 GGACATCTGTCCGCCCTCATGTACGGCAACCGCGTGTTCACCGAGTACCCCA
- 7 GGACATCGTCAAGAACCTCTGCCCGCC

8 GGCTACACCTGGGACCGCTCCTCCTGTTCGA
9 GGACGGCCCGTGTGCATCTGCAACGCCGACATCACCGTGAGCGT
10 GGAGGAGAACTGCATGTACCACGAGTCCAAGTTCTAC
11 GGC GTGAAC TCCCCGCCGACGGCCC GTGATGAAGAAGATGACCGACA ACT
12 GGGAGCCCTCCTGCGAGAAGATCATCCCCGTGCCAAGCA
13 GGGCATCTTGAAGGGCGACGTGAGCATGTACCTGCTGCTGAA
14 GGACGGTGGCCGTTGCCTGCCAGTTGACCCGTGTACAA
15 GGCCAAGTCCGTGCCCGCAAGATGCCGACT
16 GGC ACTTCATCCAGCACAAGCTGACCCGCGA
17 GGACCGCAGCGACGCCAAGAACAGAAGTGGCACCTGACCGAGCAGCCATGCCCTCC
18 GGCTCCGCCCTGCCCTGA

Target splint-RNAs for ZsGreen1 assembly (5' to 3')

1 GGGCACGGCCTGACCAAGGAGATGACCATGAA
2 GGCATGGAGGGCTGCGTGGACGGCCACAAGTT
3 GGCGTGTACCCGGCGAGGGCATCGGCTACCC
4 GGCTTCAAGGGCAAGCAGGCCATCACCTGTG
5 GGCTTGCCTTCGCCGAGGACATCTGTCCGC
6 GGCACCGAGTACCCCAAGGACATCGTCGACTA
7 GGA ACTCCTGCCCGCCGGCTACACCTGGGAC
8 GGCTCCTCCTGTTGAGGGACGGCGCCGTGTG
9 GGCATCACCGTGAGCGTGGAGGAGAACTGCAT
10 GGGAGTCCAAGTTCTACGGCGTGAAC TCCCC
11 GGAGATGACCGACA ACTGGGAGCCCTCCTGCG
12 GGCCCCGTGCCAAGCAGGGCATCTGAAGGG
13 GGGTACCTGCTGCTGAAGGACGGTGGCCGCTT
14 GGCGACACCGTGTACAAGGCCAAGTCCGTGCC
15 GGGCAAGATGCCGACTGGCACTTCATCCAGC
16 GGCAAGCTGACCCCGAGGACCGCAGCGACCG
17 GGCACGCCATGCCCTCCGGCTCCGCCCTGCC

RT-PCR primer sequences for cloning and sequencing:

ZsG-F: ATGGCCCAGTCCAAGCAC (Sigma Aldrich, MO, USA)

ZsG-R-w-6His: CTAGTGGTGATGGTGTGATGGGCAAGGCGGAGC (Sigma Aldrich, MO, USA)

RT-PCR primer sequences for amplification of T7 promoter appended ZsGreen1 gene:

T7P-ZsG-F: CAGTAATACGACTCACTATAAGTCGCCACCATGGCCCAGTCCAAGCAGC (Sigma Aldrich, MO, USA)

ZsG-R: TCAGGGCAAGGCGGAGC (Sigma Aldrich, MO, USA)

Complementary sequence to T7 RNA polymerase promoter for in vitro RNA transcription:

FAM(6-carboxyfluorescein)- CAGTAATACGACTCACTATAGG ((Integrated DNA Technologies, IA, USA)

Sequences on the microarray (5'→3'):

Note: 3' tethered on the array surface. There are 18 segmented oligos, 17 splint oligos, and 2 quality control oligos. Multiple duplicate features were made as marked. Each feature is sized 1680 **nm** x 1232**nm**.

Copy #	Sequences
1	CCACTGTTCAAAGTTACTCTGCAGGTATGGCCTTTTTTTT (QC)
2	ACTCTGCAGGTATCGGCCACTGGCAAAGTTATCCTTTTTTTT (QC)
3	TTGGTCAGGCCGTGCTGGACTGGCATGGGGCGACCTAGTGAGTCGTATTACTGTTTTTTTT (seg)
4	ACGCAGCCCTCATCGGGTACTTCATGGTCATCCTATAGTGAGTCGTATTACTGTTTTTTTT(seg)
5	TCGCCGGTATCACGAACTTGTGGCGTCTATAGTGAGTCGTATTACTGTTTTTTTT(seg)
6	TGCTTGCCCTGAAGGGTAGCCATGCCCTATAGTGAGTCGTATTACTGTTTTTTTT(seg)
7	TCGGCGAAGGGCAAGGGGCCGCCCTCCACCACGCACAGGTTGATGCCCTAGTGAGTCGTATTACTGTTTTTTTT (seg)
8	TGGGGTACTCGGTGAACACGCGGTTGCCGTACATGAAGGCGCGACAAGATGCTCTAGTGAGTCGTATTACTG TTTTTTTT(seg)
9	GCGGGGCAGGAGTTCTGAAGTAGTCGACGATGCTCTAGTGAGTCGTATTACTGTTTTTTTT(seg)
10	TCGAACAGGAAGGAGCGGTCCCAGGTGAGCCTATAGTGAGTCGTATTACTGTTTTTTTT(seg)
11	ACGCTCACGGTATGTCGGCGTTGCAGATGCACACGGCGCGTCTATAGTGAGTCGTATTACTGTTTTTTTT(seg)
12	GTAGAACTTGGACTCGTGGTACATGCAGTTCTCCTCTATAGTGAGTCGTATTACTGTTTTTTTT(seg)
13	AGTTGTCGGTCATCTCTCATCACGGGCCGTCGGCGGGAGTTACGCTATAGTGAGTCGTATTACTGTTTTTTTT TTT(seg)
14	TGCTTGGGCACGGGATGATCTCTCGCAGGAGGGCTCCCTAGTGAGTCGTATTACTGTTTTTTTT(seg)
15	TTCAGCAGCAGGTACATGCTCACGTCGCCCTCAAGATGCCCTAGTGAGTCGTATTACTGTTTTTTTT(seg)
16	TTGTACACGGTTCGAACGGCAGGCCCTAGTGAGTCGTATTACTGTTTTTTTT(seg)
17	AGTCGGGCATTTGGGGCACGGACTTGGCCTAGTGAGTCGTATTACTGTTTTTTTT(seg)
18	TCGCGGGTCAGCTGTGGATGAAGTGCCTATAGTGAGTCGTATTACTGTTTTTTTT(seg)
19	GGAGGCGATGGCGTGCTCGGTAGGTGCCACTCTGGTCTGGCGTCGCTCGGTCTAGTGAGTCGTATTACTGTTTTTTTT GTTTTTTTTT(seg)
20	TCAGGGCAAGGCGGAGCCTAGTGAGTCGTATTACTGTTTTTTTT(seg)
21	GCACGGCCTGACCAAGGAGATGACCATGAACCTATAGTGAGTCGTATTACTGTTTTTTTT(splint)
22	CATGGAGGGCTGCGTGGACGCCACAAGTCCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
23	CGTGATACCGCGAGGGCATCGGCTACCCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
24	CTTCAAGGGCAAGCAGGCCATCACCTGTGCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
25	CTTGCCTTCGCCAGGACATCTGTCGCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
26	CACCGAGTACCCCCAGGACATCGTCGACTACCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
27	AACTCCTGCCCGCCGGTACACCTGGACCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
28	CTCCTCCTGTTGAGGACGGCGCCGTGTCGCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
29	CATCACCGTGAGCGTGGAGGAGAACTGCATCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
30	GAGTCCAAGTTCTACGGCGTGAACCTCCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
31	AGATGACCGACAACGGCCCTCTGCGCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
32	CCCCGTGCCAACGCAGGGCATCTGAAGGGCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
33	GTACCTGCTGCTGAAGGACGGTGGCCGCTCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
34	CGACACCGTGTACAAGGCCAGTCGGTGCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
35	GCAAGATGCCGACTGGCACTTCATCCAGCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
36	CAAGCTGACCCCGAGGACCGCAGCGACGCCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)
37	CACGCCATGCCCTCCGGCTCCGCCTAGTGAGTCGTATTACTGTTTTTTTT(splint)

Table S1

We compare here the RNA-mediated assembly technology with other recently published gene assembly technologies. The RNA-mediated strategy drastically reduces the time and labor required for high fidelity gene synthesis from weeks to a days and no specialized equipment is needed (not including the array fabrication).

Comparison of selected strategies for gene assembly from DNA arrays

	Kosuri et al. ^[11]	Matzas et al. ^[12]	Wu et al. (this study)
1.	Production of oligonucleotide library (Agilent OLS) (proprietary synthesis, cleavage and amplification. 7 d)*	Cleavage of oligos from the arrays, and purification.	T7 RNA amplification from the array (4 h~12h)*
2.	Assembly –specific PCR amplification (2 h)*	Amplification with emulsion PCR	Buffer exchange (1 h)*
3.	Purification for size verification	Sequencing with next generation sequencer	Pyrophosphate removal (2 h)*
4.	Reamplification in 20 mL with chemically modified assembly-specific primers	Sequences selected and localized	T4 RNA ligation to produce full-length RNA transcript (40 min)*
5.	Split into 96 well plates	Bead extraction with a micro actuator from a picotiter plate to 96-well plate	Reverse transcription PCR for full-length coding gene (3 hr)*
6.	Buffer exchange (cleanup)	Amplification of DNA individually from beads. Variable biotinylated primers used to remove restriction products containing biotin residues by streptavidin matrix.	
7.	Protease digestion followed by heat inactivation	Gel purification–estimation of oligo concentration and mixing the amplicons in equimolar amounts	
8.	Protein removal	Removal of primer regions	
9.	Buffer exchange (cleanup)	Ethanol precipitated	
10.	Lambda exonuclease digestion	Polymerase cycling assembly	
11.	Buffer exchange		
12.	DpnII digestion and USER enzyme (NEB) with guide oligo		
13.	Buffer exchange (cleanup)		
14.	Polymerase cycling assembly (4 h)*		
15.	Error correction (3.5 h)*		
16.	Column reaction cleanup		
17.	Restriction digestion (2 h)*		
18.	Gel purification (2 h)*		

*estimated by the article authors.

Text

Alignment of Sanger Sequencing Data of ZsGreen1 Assemblies

ZsGreen1 gene assemblies from DNA arrays fabricated on either amorphous carbon surfaces (sequence #1 to #25) or on silanized glass surfaces (sequence #26 to #51) were Sanger sequenced (Functional Biosciences, Inc., WI, USA) and aligned with the ZsGreen1 target sequence. It should be noted sequenced nucleotides #1 to #18 correspond to the forward primer sequence (ZsG-F), and sequenced nucleotides #680 to #714 correspond to the reverse primer sequence (ZsG-R-w-6His) for ZsGreen1 RT-PCR amplification. Excluding the primer regions, 16,525 assembled nucleotides from the DNA array fabricated on the amorphous carbon surface were analyzed and 25 transitions, 3 transversions, 1 deletion, and no insertions were identified, which corresponds to an error rate of 0.1755%; whereas 17,186 assembled nucleotides from the DNA array on the silanized glass surface were analyzed and 24 transitions, no transversion, 1 deletion, and 1 insertion were identified, which corresponds to an error rate of 0.1513%. This sequence analysis of cloned constructs indicated a yield of correct constructs of approximately 40%.

Analyzing the primer sequences (character bordered), which were conventionally column synthesized from Sigma Aldrich, 1 transitions, 2 transversions, 3 deletions, and 2 insertions were identified in the ZsG-R-w-6His primer region (35 nt long; 1,785 nucleotides were analyzed; corresponds to an error rate of 0.448%) whereas no errors were found in the short ZsG-F primer region (18 nt long).

	10	20	30	40	50	60	70
ZsGreen1
	ATGGCCCAGTCCAAGCACGGCCTGACCAAGGAGATGACCATGAAGTACCGCATGGAGGGCTGCGTGGACG						
01A_1.ab1
01B_2.ab1
01C_3.ab1
01D_4.ab1
01E_5.ab1
01F_6.ab1
01G_7.ab1	A
01H_8.ab1
02A_9.ab1
02B_10.ab1
02C_11.ab1
02D_12.ab1
02E_13.ab1	G
02F_14.ab1
02G_15.ab1
02H_16.ab1
03A_17.ab1
03B_18.ab1
03C_19.ab1
03D_20.ab1
03E_21.ab1
03F_22.ab1	G
03G_23.ab1	G
03H_24.ab1
04A_25.ab1

	80	90	100	110	120	130	140
ZsGreen1
	GCCACAAGTTCTGTGATCACCGGCCAGGGCATCGGCTACCCCTTCAAGGGCAAGCAGGCCATCAACCTGTG						
01A_1.ab1
01B_2.ab1
01C_3.ab1
01D_4.ab1
01E_5.ab1	G
01F_6.ab1
01G_7.ab1
01H_8.ab1
02A_9.ab1
02B_10.ab1
02C_11.ab1
02D_12.ab1
02E_13.ab1
02F_14.ab1
02G_15.ab1
02H_16.ab1
03A_17.ab1
03B_18.ab1
03C_19.ab1
03D_20.ab1
03E_21.ab1
03F_22.ab1
03G_23.ab1
03H_24.ab1
04A_25.ab1

	150	160	170	180	190	200	210
ZsGreen1	CGTGGTGGAGGGCGGCCCCTTGCCCTTCGCCGAGGACATCTTGTCCGCCCTTCATGTAACGGCAACCGC						
01A_1.ab1
01B_2.ab1
01C_3.ab1
01D_4.ab1
01E_5.ab1
01F_6.ab1
01G_7.ab1
01H_8.ab1
02A_9.ab1
02B_10.ab1
02C_11.ab1
02D_12.ab1
02E_13.ab1
02F_14.ab1
02G_15.ab1
02H_16.ab1
03A_17.ab1	A.
03B_18.ab1
03C_19.ab1
03D_20.ab1
03E_21.ab1	C.
03F_22.ab1
03G_23.ab1
03H_24.ab1
04A_25.ab1

	220	230	240	250	260	270	280
ZsGreen1	GTTGTTACCGAGTACCCCCAGGACATCGTCGACTACTTCAAGAACTCCTGCCCGCTACACCTGGG						
01A_1.ab1
01B_2.ab1
01C_3.ab1
01D_4.ab1
01E_5.ab1
01F_6.ab1
01G_7.ab1
01H_8.ab1
02A_9.ab1
02B_10.ab1
02C_11.ab1
02D_12.ab1
02E_13.ab1
02F_14.ab1
02G_15.ab1
02H_16.ab1
03A_17.ab1
03B_18.ab1	T
03C_19.ab1	T
03D_20.ab1
03E_21.ab1
03F_22.ab1
03G_23.ab1
03H_24.ab1
04A_25.ab1

	430	440	450	460	470	480	490
ZsGreen1	ATGACCGACA	ACTGGGAGCC	CTCCTGC	GAGAAGATCA	TCCCCGTGCC	AAGCAGGGCAT	CTTGAAGGGCG
01A_1.ab1
01B_2.ab1
01C_3.ab1
01D_4.ab1
01E_5.ab1
01F_6.ab1
01G_7.ab1
01H_8.ab1
02A_9.ab1
02B_10.ab1
02C_11.ab1
02D_12.ab1
02E_13.ab1
02F_14.ab1
02G_15.ab1
02H_16.ab1
03A_17.ab1
03B_18.ab1
03C_19.ab1
03D_20.ab1
03E_21.ab1
03F_22.ab1
03G_23.ab1
03H_24.ab1
04A_25.ab1

	500	510	520	530	540	550	560
ZsGreen1	ACGTGAGCATG	TACCTGCTG	CTGAAGGACGG	TGGCCGCTTG	CGCTGCCAGTT	CGACACCGTG	TACAAGGC
01A_1.ab1
01B_2.ab1
01C_3.ab1
01D_4.ab1
01E_5.ab1
01F_6.ab1
01G_7.ab1
01H_8.ab1
02A_9.ab1
02B_10.ab1
02C_11.ab1
02D_12.ab1
02E_13.ab1
02F_14.ab1
02G_15.ab1
02H_16.ab1
03A_17.ab1
03B_18.ab1
03C_19.ab1
03D_20.ab1
03E_21.ab1
03F_22.ab1
03G_23.ab1
03H_24.ab1
04A_25.ab1

	570	580	590	600	610	620	630
ZsGreen1
01A_1.ab1
01B_2.ab1
01C_3.ab1	A
01D_4.ab1
01E_5.ab1
01F_6.ab1
01G_7.ab1
01H_8.ab1	G
02A_9.ab1
02B_10.ab1	T
02C_11.ab1
02D_12.ab1
02E_13.ab1
02F_14.ab1
02G_15.ab1
02H_16.ab1
03A_17.ab1
03B_18.ab1
03C_19.ab1
03D_20.ab1
03E_21.ab1
03F_22.ab1
03G_23.ab1
03H_24.ab1
04A_25.ab1

	640	650	660	670	680	690	700
ZsGreen1
01A_1.ab1
01B_2.ab1
01C_3.ab1
01D_4.ab1
01E_5.ab1
01F_6.ab1
01G_7.ab1
01H_8.ab1
02A_9.ab1
02B_10.ab1
02C_11.ab1
02D_12.ab1
02E_13.ab1
02F_14.ab1	T
02G_15.ab1
02H_16.ab1
03A_17.ab1
03B_18.ab1
03C_19.ab1
03D_20.ab1
03E_21.ab1
03F_22.ab1
03G_23.ab1
03H_24.ab1
04A_25.ab1

710

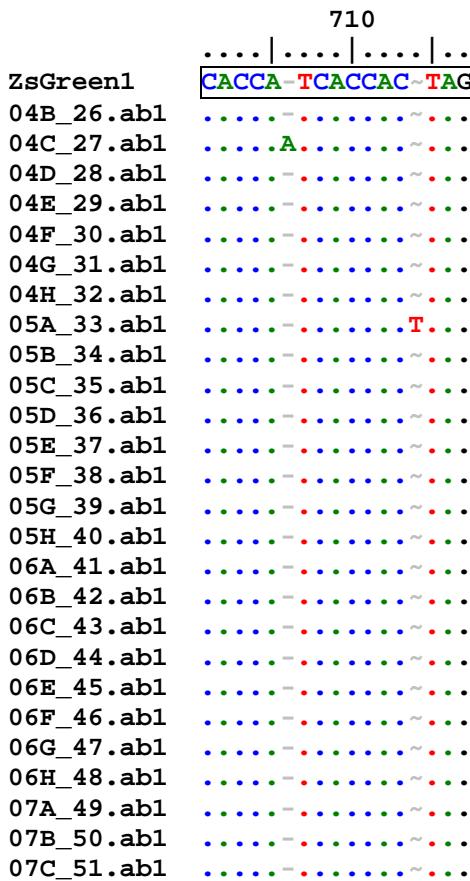
ZsGreen1	ACCATCACCAC TAG
01A_1.ab1
01B_2.ab1
01C_3.ab1
01D_4.ab1
01E_5.ab1
01F_6.ab1
01G_7.ab1
01H_8.ab1
02A_9.ab1
02B_10.ab1
02C_11.ab1
02D_12.ab1
02E_13.ab1
02F_14.ab1
02G_15.ab1
02H_16.ab1
03A_17.ab1
03B_18.ab1
03C_19.ab1
03D_20.ab1
03E_21.ab1
03F_22.ab1
03G_23.ab1
03H_24.ab1
04A_25.ab1

	10	20	30	40	50	60	70
ZsGreen1	ATGGCCCAGTCCAAGCAC	GGCCTGACCAAGGAGATGACCATGAAGTACCGCATGGAGGGCTGCGTGGACG					
04B_26.ab1
04C_27.ab1
04D_28.ab1
04E_29.ab1
04F_30.ab1
04G_31.ab1
04H_32.ab1
05A_33.ab1
05B_34.ab1
05C_35.ab1
05D_36.ab1
05E_37.ab1
05F_38.ab1
05G_39.ab1
05H_40.ab1
06A_41.ab1
06B_42.ab1
06C_43.ab1
06D_44.ab1
06E_45.ab1
06F_46.ab1
06G_47.ab1
06H_48.ab1
07A_49.ab1
07B_50.ab1
07C_51.ab1

	80	90	100	110	120	130	140
ZsGreen1	GCCACAAGTTCTGTGATCACC	GGCGAGGGCATCGGCTAACCC	TTCAAGGGCAAGCAGGCCAT	CAACCTGTG			
04B_26.ab1
04C_27.ab1
04D_28.ab1
04E_29.ab1
04F_30.ab1
04G_31.ab1
04H_32.ab1
05A_33.ab1
05B_34.ab1
05C_35.ab1
05D_36.ab1
05E_37.ab1
05F_38.ab1	A.
05G_39.ab1
05H_40.ab1
06A_41.ab1	.	A.
06B_42.ab1
06C_43.ab1
06D_44.ab1
06E_45.ab1
06F_46.ab1
06G_47.ab1
06H_48.ab1
07A_49.ab1
07B_50.ab1
07C_51.ab1

	430	440	450	460	470	480	490
ZsGreen1
04B_26.ab1	G	A	T	G	A	C	G
04C_27.ab1	A	T	G	A	C	G	G
04D_28.ab1	T	C	C	T	C	C	C
04E_29.ab1	C	C	A	G	A	G	G
04F_30.ab1	C	C	G	A	G	G	G
04G_31.ab1	C	C	G	A	G	G	G
04H_32.ab1	C	C	G	A	G	G	G
05A_33.ab1	C	C	G	A	G	G	G
05B_34.ab1	C	C	G	A	G	G	G
05C_35.ab1	C	C	G	A	G	G	G
05D_36.ab1	C	C	G	A	G	G	G
05E_37.ab1	C	C	G	A	G	G	G
05F_38.ab1	C	C	G	A	G	G	G
05G_39.ab1	C	C	G	A	G	G	G
05H_40.ab1	C	C	G	A	G	G	G
06A_41.ab1	C	C	G	A	G	G	G
06B_42.ab1	C	C	G	A	G	G	G
06C_43.ab1	C	C	G	A	G	G	G
06D_44.ab1	C	C	G	A	G	G	G
06E_45.ab1	C	C	G	A	G	G	G
06F_46.ab1	C	C	G	A	G	G	G
06G_47.ab1	C	C	G	A	G	G	G
06H_48.ab1	C	C	G	A	G	G	G
07A_49.ab1	C	C	G	A	G	G	G
07B_50.ab1	C	C	G	A	G	G	G
07C_51.ab1	C	C	G	A	G	G	G

	500	510	520	530	540	550	560
ZsGreen1	GACGTGAGCATGTACCTGCTGCTGAAGGACGGTGGCCGCTTGCGCTGCCAGTTCGACACCGTGTACAAGG						
04B_26.ab1	
04C_27.ab1	
04D_28.ab1	
04E_29.ab1	
04F_30.ab1	
04G_31.ab1	
04H_32.ab1	
05A_33.ab1	
05B_34.ab1	
05C_35.ab1	
05D_36.ab1	
05E_37.ab1	
05F_38.ab1	
05G_39.ab1	
05H_40.ab1	
06A_41.ab1	
06B_42.ab1	
06C_43.ab1	
06D_44.ab1	
06E_45.ab1	
06F_46.ab1	
06G_47.ab1	
06H_48.ab1	
07A_49.ab1	
07B_50.ab1	
07C_51.ab1	
			A				



References:

- [1] B. L. Nilsson, M. B. Soellner, R. T. Raines, *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 91-118.
- [2] K. Itakura, T. Hirose, R. Crea, A. D. Riggs, H. L. Heyneker, F. Bolivar, H. W. Boyer, *Science* **1977**, *198*, 1056-1063.
- [3] W. P. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan, H. L. Heyneker, *Gene* **1995**, *164*, 49-53.
- [4] M. Kozak, *Nucleic Acids Res* **1987**, *15*, 8125-8148.
- [5] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res* **1987**, *15*, 8783-8798.
- [6] C. T. Martin, D. K. Muller, J. E. Coleman, *Biochemistry* **1988**, *27*, 3966-3974.
- [7] Z. Guo, R. A. Guilfoyle, A. J. Thiel, R. F. Wang, L. M. Smith, *Nucleic Acids Res* **1994**, *22*, 5456-5465.
- [8] a) S. Singh-Gasson, R. D. Green, Y. J. Yue, C. Nelson, F. Blattner, M. R. Sussman, F. Cerrina, *Nat. Biotechnol.* **1999**, *17*, 974-978; b) M. F. Phillips, M. R. Lockett, M. J. Rodesch, M. R. Shortreed, F. Cerrina, L. M. Smith, *Nucleic Acids Res* **2008**, *36*, e7.
- [9] a) M. R. Lockett, L. M. Smith, *Anal Chem* **2009**, *81*, 6429-6437; b) M. R. Lockett, S. C. Weibel, M. F. Phillips, M. R. Shortreed, B. Sun, R. M. Corn, R. J. Hamers, F. Cerrina, L. M. Smith, *Journal of the American Chemical Society* **2008**, *130*, 8611-8613.
- [10] J. F. Milligan, O. C. Uhlenbeck, *Methods in Enzymology* **1989**, *180*, 51-62.
- [11] a) K. E. Richmond, M. H. Li, M. J. Rodesch, M. Patel, A. M. Lowe, C. Kim, L. L. Chu, N. Venkataraman, S. F. Flickinger, J. Kayser, P. J. Belshaw, M. R. Sussman, F. Cerrina, *Nucleic Acids Res* **2004**, *32*, 5011-5018; b) J. D. Tian, H. Gong, N. J. Sheng, X. C. Zhou, E. Gulari, X. L. Gao, G. Church, *Nature* **2004**, *432*, 1050-1054; c) A. S. Xiong, Q. H. Yao, R. H. Peng, X. Li, H. Q. Fan, Z. M. Cheng, Y. Li, *Nucleic Acids Res* **2004**, *32*, e98; d) C. Kim, J. Kayser, K. Richmond, M. Rodesch, B. Binkowski, L. Chu, M. Li, K. Heinrich, S. Blair, P. Belshaw, M. Sussman, F. Cerrina, *Microelectron. Eng.* **2006**, *83*, 1613-1616; e) S. Kosuri, N. Eroshenko, E. M. LeProust, M. Super, J. Way, J. B. Li, G. M. Church, *Nat. Biotechnol.* **2010**, *28*, 1295-U1108; f) M. Matzas, P. F. Stahler, N. Kefer, N. Siebelt, V. Boisguerin, J. T. Leonard, A. Keller, C. F. Stahler, P. Haberle, B. Gharizadeh, F. Babrzadeh, G. M. Church, *Nat. Biotechnol.* **2010**, *28*, 1291-U1101.
- [12] E. Hochuli, W. Bannwarth, H. Dobeli, R. Gentz, D. Stuber, *Bio-Technol* **1988**, *6*, 1321-1325.
- [13] a) J. Y. Quan, I. Saaem, N. Tang, S. M. Ma, N. Negre, H. Gong, K. P. White, J. D. Tian, *Nat. Biotechnol.* **2011**, *29*, 449-U239; b) A. Y. Borovkov, A. V. Loskutov, M. D. Robida, K. M. Day, J. A. Cano, T. Le Olson, H. Patel, K. Brown, P. D. Hunter, K. F. Sykes, *Nucleic Acids Res* **2010**, *38*, e180.
- [14] A. S. Xiong, R. H. Peng, J. Zhuang, J. G. Liu, F. Gao, J. M. Chen, Z. M. Cheng, Q. H. Yao, *Biotechnol. Adv.* **2008**, *26*, 121-134.
- [15] M. A. Cleary, K. Kilian, Y. Q. Wang, J. Bradshaw, G. Cavet, W. Ge, A. Kulkarni, P. J. Paddison, K. Chang, N. Sheth, E. Leproust, E. M. Coffey, J. Burchard, W. R. McCombie, P. Linsley, G. J. Hannon, *Nat. Methods* **2004**, *1*, 241-248.
- [16] a) X. Gao, B. L. Gaffney, M. Senior, R. R. Riddle, R. A. Jones, *Nucleic Acids Res* **1985**, *13*, 573-584; b) R. T. Pon, M. J. Damha, K. K. Ogilvie, *Nucleic Acids Res* **1985**, *13*, 6447-6465; c) R. T. Pon, N. Usman, M. J. Damha, K. K. Ogilvie, *Nucleic Acids Res* **1986**, *14*, 6453-6470; d) S. Crippa, P. Digennaro, R. Lucini, M. Orlandi, B. Rindone, *Gazzetta Chimica Italiana* **1993**, *123*, 197-203.
- [17] All of the RNA segments employed for the assembly of ZsGreen1 have higher Gibbs free energy (less stable) for secondary structure folding than for hybridization to their complementary splints.
- [18] K. L. Prime, G. M. Whitesides, *Journal of the American Chemical Society* **1993**, *115*, 10714-10721.
- [19] A. C. Pease, D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes, S. P. A. Fodor, *P Natl Acad Sci USA* **1994**, *91*, 5022-5026.
- [20] G. H. McGall, A. D. Barone, M. Diggelmann, S. P. A. Fodor, E. Gentalen, N. Ngo, *J Am Chem Soc* **1997**, *119*, 5081-5090.
- [21] S. Walbert, W. Pfleiderer, U. E. Steiner, *Helvetica Chimica Acta* **2001**, *84*, 1601-1611.