Contract No.: HR0011-13-C-0073

Institution: SYNTHETIC GENOMICS, INC.

Title: DIGITAL BIOLOGICAL CONVERTER

MILESTONE ACCEPTANCE REPORT

Date of Report: 03 January 2014
DARPA Program Manager: Alicia Jackson
Technical POC: Barbara Pica

Total Dollar Value: $998,847
Start Date: JUNE 5, 2013
End Date: JUNE 4, 2014
DARPA Contracting Officer: Michael Blackstone

Submitted by:
HEATHER GOUVIS
Synthetic Genomics, Inc.
11149 NORTH TORREY PINES RD
LA JOLLA, CA 92037
Telephone: 858.433.2281
Fax: 858.754.2988
Email: hgouvis@syntheticgenomics.com

Security Classification – Unclassified

DISTRIBUTION A. Approved for public release: distribution unlimited.

This material is based upon work supported by DARPA under contract HR0011-13-C-0073

One report to:

Alicia Jackson
Defense Advanced Research Projects Agency
Alicia.Jackson@darpa.mil

Marc Salit
Contracting Officer Representative (COR)
Email: salit@nist.gov

Michael Blackstone
DARPA Contracting Officer
Michael.Blackstone@darpa.mil

DARPA/MTO (Technical Office)
ADPM, Barbara Pica
MTO-ADPM@darpa.mil

Tom Armel
SETA
thomas.armel.ctr@darpa.mil

Natalie Kent
Administrative SETA
Natalie.Kent.ctr@darpa.mil

1. Executive Summary

SGI has previously reported completion of the first three milestones. In this report we describe the successful demonstration of the fourth milestone—fully automated generation of protein starting from a DNA sequence that codes for that protein. In our example, we produced green fluorescence protein (GFP). We entered a 1.5-kb gene sequence for GFP into SGI's Archetype® that automatically generated a file containing the overlapping oligonucleotides necessary to produce the 1.5-kb gene fragment and initiated the start of the Digital-Biological Converter (DBC) run. Forty-eight overlapping ~64-base oligonucleotide oligonucleotides were synthesized, post-processed, and then pooled. The pooled oligonucleotides were then assembled and the resulting GFP gene fragment was amplified, error-corrected, and then re-amplified. This synthetic GFP gene product was then incubated in a cell-free extract and the synthetic GFP sequence was translated into protein. Multiple reactions were carried out in parallel and as much as 37 µg/mL GFP protein was produced.

This material is based upon work supported by DARPA under contract HR0011-13-C-0073

2. Milestone Summary

At SGI, an integrated DBC prototype has been engineered which synthesizes DNA starting from an oligo design file that is transmitted from SGI's Archetype®. The instrument then automatically synthesized and pooled the oligos, assembled and error corrected the resulting DNA, and translated the DNA to produce 37μg/mL of GFP protein. This was accomplished by the close collaboration of three organizations: Synthetic Genomics, BioAutomation, and Hudson Robotics. Two major additions were added to the DBC controller software:

1) The addition of incubations and pipetting routines to transcribe and translate the DNA template produced by the DBC into protein.
2) A file watch capability to automatically launch when the instrument is in the ready state and an Archetype® file arrives in a specific target directory. For full integration, our onsite bioinformatics engineers altered the Archetype® software such that any oligo design file generated with the identifier "DBC" would be transferred to a target folder on our network to launch the DBC. We opted for network transfer of the file since email is inherently insecure.

The total run time for the configuration of the DBC to produce GFP was 36h and 25m, which included a 15h and 45m synthesis of a full plate of oligos. Total deprotection was 6h 48m due to a 2.5h base incubation step and 4h of positive air pressure evaporation to reduce the base concentration for subsequent steps. Pooling was 26m followed by assembly and protein production which clocked at 13h 27m including an oligo assembly reaction, an error correction incubation step, a recovery PCR and a further incubation to synthesize GFP.

This material is based upon work supported by DARPA under contract HR0011-13-C-0073

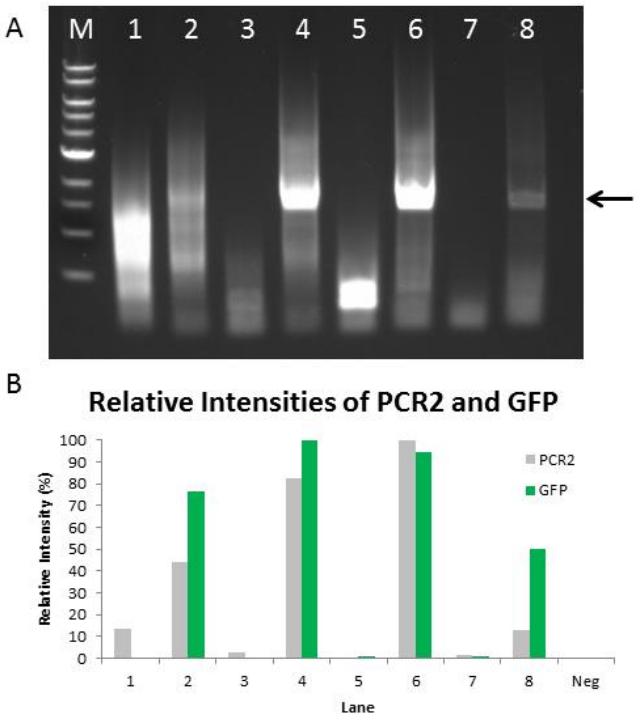

Fig 1. DNA template bands and the resulting GFP fluorescence. A) DNA templates (1.5kb) of various concentrations and yields were produced after dilution of the starting oligo pool. B) Relative intensities of the DNA templates and their corresponding GFP signals by fluorescent spectroscopy. Intensities were normalized to 100%.

Milestone 4 Tasks:

a. Evaluation of cell free protein expression kit(s)

- Kits were purchased from Promega, Invitrogen, and NEB and used to produce GFP from the 1.5kb template used in Milestone3. Although several cell-free transcription/translation kits produced GFP, the NEB PureExpress kit was chosen as it was stable at 4°C for up to 40h.

b. Evaluation of *in vitro* protein expression methods (PCR template)

- Variations of the 1.5kb GFP template from Milestone3 were made and tested for GFP yield. The overall design was a 5' end homologous to a standard end primer, a 660bp stuffer sequence to increase the total length to 1.5kb, a T7-promoter followed by the open reading frame for GFP, a series of in-frame stop codons and a T7-terminator ending with a terminal primer sequence for amplification. Alterations in the translation-initiation and termination sites and the addition of a T7-terminator sequence improved the yield by as much as 10-fold over the original template.

c. Process optimization and protein quantitation methods used

- Fluorescent spectroscopy protocols were developed to measure the activity of GFP. SDS protein gels were standardized to resolve the protein band for GFP to allow densitometry-assisted quantification of the bands for concentration. Additionally, any truncated proteins would be observable as degradation products (not observed).

This material is based upon work supported by DARPA under contract HR0011-13-C-0073

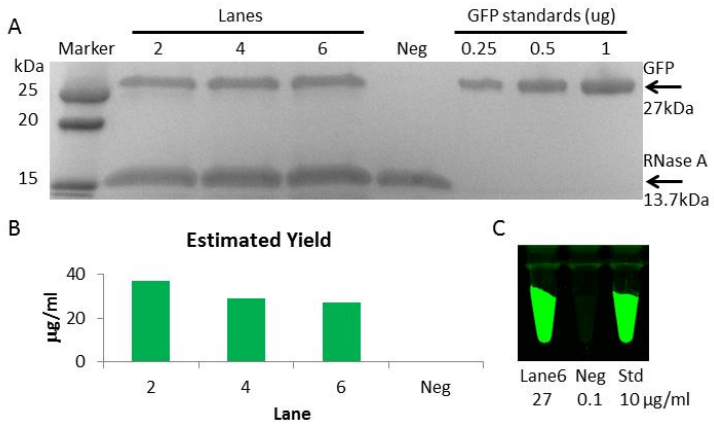

Fig 2. SDS Protein gel showing yield of full length GFP. A) 50µl of translation reactions from Lanes 2, 4, and 6 were loaded on an SDS gel along with purified GFP standards of known concentration and non-template control. B) Estimated yield of GFP by densitometry. Intensity of each GFP band was normalized to RNase A in the translation reaction as a loading control and yield was estimated by standard curve. Yields were 37, 29, and 27 µg/ml for Lanes 2, 4, and 6, respectively. C) A fluorescent image of the GFP from Lane 6 compared to negative and positive controls.

d. DNA design and assembly of published DNA sequence for GFP

- The GFP template was derived from a previously published GFP variant, Genbank Accession AF325903. Oligonucleotide assembly was robust. Surprisingly, a standard primer for pUC19 vector consistently misprimed in the GFP template, even with low identity. We eventually selected other end primers to avoid the mispriming.

e. Additional integration of automated capability to execute protein expression protocols to DBC robotics

- Additions were made to the assembly routines for pipetting and incubation steps necessary to produce protein. The NEB PURExpress translation system was selected since it performed well, even after incubation on the DBC instrument up to 40h. Besides the addition of protein production and automatic launch, alterations were made to the deprotection routine to reduce the volume and the base content, thus improving the oligonucleotide assembly process and the yield of the GFP template. The GFP protein yield was dependent on the amount of template added to the translation-mix. A maximum amount of template was determined that could be added to the translation reaction without causing inhibition due buffer and polymerase carry over.

This material is based upon work supported by DARPA under contract HR0011-13-C-0073

3. Issues:

- Run time greater than a day

  The entire run time to reach GFP protein production was over 36h. We had to systematically identify assembly and translation components that would still perform after such a long incubation in the 4C bath on the DBC.

- Base carry over from deprotection into oligonucleotide assembly

  The original protocols for deprotection of oligos from the BioAutomation's Mermade 192E involved heating the oligos to 70°C for 2-3h under positive pressure to dryness following deprotection. The DBC does not have a heated block capability, hence, 2-4h of negative pressure was used instead. The stringent requirement for a high concentration template necessitated an improved oligonucleotide assembly by reducing the base carried over from deprotection into assembly as much as possible. To achieve this, we substituted positive pressure for 4h as opposed to the negative pressure in the original protocol. We observed that around half of the AMA solution containing the deprotected oligos evaporated over this time at room temperature. This modification was sufficient to improve the yield of template since the base components of AMA are poorly soluble in water and evaporate more rapidly than the aqueous solution.

4. Technology Transitions: None to report at this time.

This material is based upon work supported by DARPA under contract HR0011-13-C-0073