Accurate High-throughput Gene Synthesis Using Programmable Microfluidic DNA Microchips

Jingdong Tian*, Hui Gong*, Nijing Sheng**, Xiaochuan Zhou**, Erdogan Gulari***, Xiaolian Gao**, and George Church*

*Harvard Medical School, Boston, MA, USA, jtianyh@yahoo.com
**University of Houston, Houston, TX, USA
***University of Michigan, Ann Arbor, Michigan, USA

ABSTRACT

Testing the deluge of hypotheses flowing from genomics and systems biology demands accurate and cost-effective gene and genome synthesis. Here we describe a microchip-based technology for multiplex gene synthesis[1]. Pools of thousands of "construction" oligonucleotides and tagged complementary "selection" oligonucleotides are synthesized on photo-programmable microfluidic chips, released, amplified, and hybridization-selected to reduce synthesis errors. A one-step polymerase assembly multiplexing (PAM) reaction assembles these into multiple genes. This technology is promising for high-throughput gene and genome syntheses and should have utility for synthetic biology in general.

Keywords: microfluidics, gene synthesis, DNA microchip, microfabrication, high-throughput

1 INTRODUCTION

The advance of biochemical analyses (sequencing, microarrays, and proteomics) has generated vast amounts of data and computational biology has leveraged these into a huge number of hypotheses. To test these and extend them to useful designed systems we must remedy the bottleneck in constructing new genetic elements, genetic pathways, and engineered cells. These represent great challenges and payoffs for the emerging field of synthetic biology.

In principle, we can create a useful variety of molecules, cellular and cell-free systems given a sufficient supply of custom genes and genomes. However, current methods for generating even basic oligonucleotides are expensive and have very high levels of errors (deletions at a rate of 1/100 bases and mismatches and insertions at about 1/400 bases). As a result, gene or genome synthesis from oligonucleotides is both expensive and error-prone. Correcting errors by clone sequencing and mutagenesis methods further increases the degree of labor and total cost. In principle, the cost of oligonucleotide synthesis can be reduced by massively parallel, custom syntheses on microchips. However, current chips have very low surface areas and hence low amounts of oligonucleotides can be produced. When released into solution, each oligo is present in very low concentrations, insufficient to drive bimolecular priming reactions efficiently.

2 RESULTS

A potential solution to this scale problem would be amplification of the oligonucleotides obtained from the chips from roughly as little as 10^5 (or 10^9 for low density arrays) up to 10^9 (or 10^{12}) molecules of each sequence, thereby permitting subsequent selection and assembly steps.

In this general method, shown in Fig. 1a, oligonucleotides flanked by short, generic adapter sequences are synthesized on a programmable microchip. This generates a pool of 10^2 to 10^5 different oligonucleotides that can be released from the chips by chemical or enzymatic treatment. Released oligonucleotides are amplified by PCR using primers carrying type-IIS restriction enzyme recognition sites. Digestion of the PCR products with the corresponding restriction enzyme(s) yields sufficient amounts of unadulterated oligonucleotide sequences to be used for gene or genome assembly.

We first demonstrated the feasibility of this approach with Xeotron 4K photo-programmable microfluidic microarrays. To monitor oligonucleotide synthesis and cleavage from the chip, the 5' end of the oligonucleotides were coupled with fluorescein. The chip was scanned with a microarray scanner before and after cleavage, and the images are shown in Fig. 1b. The cleaved oligonucleotides were hybridized onto a "QA-chip" synthesized with complementary oligonucleotide sequences (Fig. 1c). These results demonstrated that individual oligonucleotides were synthesized and nearly completely released from the chip in quantities that can be measured by a QA-chip hybridization process. The typical yield of oligonucleotide released from each chamber of the 4K chip is about 5 fmoles as determined by quantitative PCR. Using primers that anneal specifically to the generic adapters flanking the

oligonucleotide sequences, PCRs were carried out that would amplify the oligonucleotides more than a million-fold.

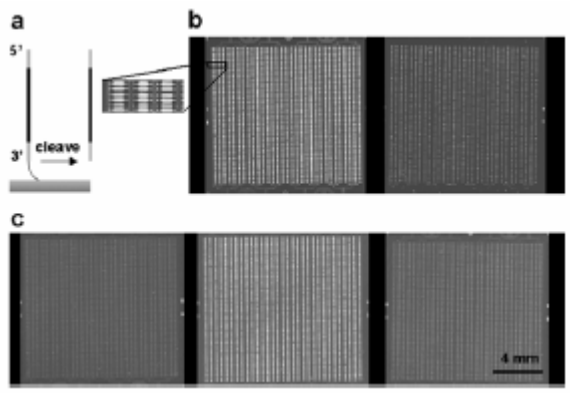

Figure 2. Preparation of free oligonucleotides from custom microarray. a, Diagram of synthesis and cleavage of a PCR-amplifiable oligonucleotide from microchip surface. The portion of the oligonucleotide used for gene construction is in black, PCR primer-adapters are in gray. b, Synthesis and cleavage of oligonucleotides from a Xeotron/Atactic 4K photo-programmable microfluidic microchip. Left, fluorescent scanning micrograph of an oligonucleotide array before cleavage. Insert, details of microfluidic chambers and connecting channels. Right, array after cleavage. c, Hybridization of released FAM-labeled oligonucleotides to a QA-chip. Left, before hybridization; middle, after hybridization; right, after stripping of hybridized oligonucleotides.

Mutations incurred during oligonucleotide synthesis are a major source of errors in assembled DNA molecules and are costly and difficult to eradicate[2-3]. We developed a simple, stringent hybridization-based method to remove oligonucleotides with such mutations. To select against mutations in gene construction oligonucleotides, these oligonucleotides were hybridized sequentially to two pools of bead-immobilized short complementary selection oligonucleotides that together span the entire length of the construction oligonucleotides. All selection oligonucleotides were designed to have nearly identical melting temperatures with varying lengths. Under proper hybridization conditions, imperfect pairs between selection and construction oligonucleotides due to base-mismatch or deletion have lower melting temperatures and are unstable. After the hybridization-wash-elution cycles, oligonucleotides with perfect matching sequence to the selection oligonucleotides are preferentially retained and enriched. Digestion of the PCR products with type-IIS restriction enzymes removed the generic primer sequences from both ends of the oligonucleotides. In these experiments the amplification tags are removed just before selection. However, if the digestion were deferred, the oligonucleotides could be re-amplified by PCR and subjected to further rounds of hybridization selection. Since the chance of complementary mutations that occur at matching positions on construction and selection oligonucleotides is miniscule, in principle most oligonucleotides with mutations can be eliminated by this selection procedure.

Like construction oligonucleotides, selection oligonucleotides were also synthesized and released from programmable microarrays. Selection oligonucleotides with arms were amplified by PCR, and the strand complementary to the gene construction oligonucleotide was labeled with biotin at the 5'-end and selectively immobilized on streptavidin beads. The unlabeled strand was denatured and removed.

Such oligonucleotides are suitable for gene assembly. To facilitate automation, we developed a single-step polymerase assembly multiplexing (PAM) reaction for multiple gene syntheses from a single pool of oligonucleotides. Single fragment assembly methods have traditionally used two or three steps (ligation, assembly and PCR)[2-4]. For PAM, gene-flanking primer pairs were added to the same pool of gene-construction oligonucleotides at a higher concentration, together with thermostable polymerase and dNTPs. Extension of overlapping oligonucleotides and subsequent amplification of multiple full-length genes were accomplished in a closed-tube, one-step reaction using a thermal cycler. Different generic adaptor sequences can be incorporated into the ends of each gene or gene set, and a set complementary adaptor-primer pairs can be pre-synthesized to avoid the cost of synthesizing gene-specific PAM primer pairs and to facilitate automation (*e.g.* 96 or 384 generic adaptors to match standard multi-well plates).

To determine the efficiency of the hybridization-selection method to eliminate mismatch mutations, we constructed genes using the same pool of chip-synthesized oligonucleotides purified in three different ways: 1) unpurified, 2) PAGE-purified, and 3) hybridization-purified. These genes were cloned and random clones from each category were sequenced in both directions to determine error types and rates in each category. It was found that genes synthesized with unpurified oligonucleotides have the highest error rates, ~1/160 bp; the method of gene assembly, either with ligation or PAM, made little difference. PAGE-purification of oligonucleotides helped reduce the error rate to ~1/450 bp, mainly through removal of deletion mutations. This rate is comparable to figures reported by other groups using PAGE-purification[4,5]. With hybridization selection, the error rate was further reduced to approximately 1/1,394 bp.

As an example of the usefulness of the technology to large scale synthetic biology projects, we redesigned and

synthesized from a microchip codon-altered versions of the 21 protein-encoding genes that constitute the *E. coli* small ribosomal subunit (Fig. 2).

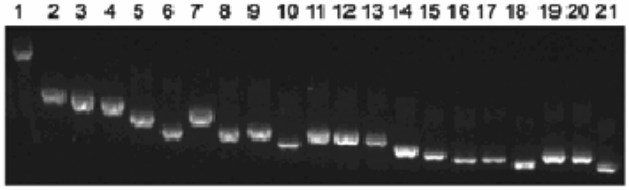

Figure 2. 21 engineered ribosomal protein genes constructed using PAM reactions.

3 DISCUSION

In this demonstration, a small fraction of a chip's synthetic capacity was utilized. Chip-based massively parallel oligonucleotide synthesis could reduce the cost of oligonucleotide synthesis to 20 kbp per $ (from 9 bp per $), depending on the type of chip used and the number of oligonucleotides on a chip. To achieve few or no errors in megabase-scale assemblies, the progress here on mismatch errors must be extended. In addition to physical methods described here, we and others are employing protein-based methods (*e.g.* using MutS or MutHLS) already shown capable of exceeding by ten-fold the fidelity of DNA polymerases ($6x10^{-6}$ per base per cycle)[5-6]. This could significantly reduce the cost spent on sequencing and potentially eliminate the need for error correction by directed-mutagenesis methods. Ultra-low-cost DNA sequencing[7] could also have a large impact if properly integrated. The ongoing next stage in testing the limits to simultaneous synthesis and assembly is 95K-382K oligonucleotides per $700 chip from Nimblegen (2 to 18 Mb). Overcoming challenges of repeated sequences may require hierarchical assembly, assembly focused on only the unique ends, and/or multiplex size selections on assembly products. Geometrical constraints to force a certain order of assembly illustrates another potential strength of microfluidic syntheses. Improving error rates enables a variety of applications without cloning or sequencing.

Accurate, low-cost multiplex gene syntheses will be a powerful tool for synthetic biology and complex nanostructures in general. For example, this technology is currently enabling us to improve and test components needed for the synthesis of ribosomes in vitro. The rapid prototyping of individual genes on short linear templates has greatly aided debugging of this complex system. Our ability to remap a whole set of genes from the *Mycoplasma*[14] codons to those of *E. coli* (*e.g.* elimination of the UGA "stop" codon, and shift from 25% to 51% G+C) will help us to calibrate proteomics experiments, test de novo protein designs and to identify new biochemical activities.

4 METHODS

4.1 Design of sequences

Gene and oligonucleotide sequences were designed using a Java program called CAD-PAM (computer-aided design polymerase assembly multiplexing) to be described in detail elsewhere (J.T., H.G., & G.C., manuscript in preparation). Basically, constraints on the amino acid sequences, codon usage, mRNA secondary structure, and restriction enzymes used to release the construction oligonucleotides were used by CAD-PAM to create nearly optimal overlapping sets of n-mers (typically 50-mers) construction oligomers and shorter selection oligomers (typically 26-mers). The melting temperatures (T_m) of overlapping regions between adjacent gene construction oligonucleotides or between construction and selection oligonucleotides were equalized, respectively. The selection oligonucleotides were padded with extra "A"s to keep oligomer length constant (70-mers) for optional size selection (not used for typical PAM). T_m values were calculated by nearest neighbor method. Codons can be fixed or altered to allow expression improvements.

4.2 Chip synthesis, amplification and selection of oligonucleotides

Oligonucleotides were synthesized on photo-programmable microfluidic microchips with a phosphate at the 5' end and its 3'-end coupling to the 3'-OH of a uracil residue. (Gao *et al.*, in preparation). After synthesis, the oligonucleotides were cleaved either with RNase A or by ammonium hydroxide treatment used for deprotection as in standard oligonucleotide syntheses followed by precipitation. Gene construction oligonucleotides PCR amplified with 20-mers (initially complementary to the terminal 10 bases) were digested with type-IIS restriction enzymes *Bsa*I and *Bse*RI. Immobilization of biotin-labeled selection oligonucleotides on magnetic streptavidin beads (Dynal) and removal of nonbiotinylated strand were done as described[16]. Construction oligonucleotides were denatured at 95ºC for 3 min and hybridized to selection oligonucleotides in hybridization buffer (5x SSPET, 50% formamide, 0.2 mg ml^{-1} BSA) for 14-16 hr at 42 °C on a rotor. Beads were washed three times with 0.5 x SSPET and three times with wash buffer (20 mM Tris-HCl, pH 7.0, 5 mM EDTA, 4 mM NaCl) at room temperature. The construction oligonucleotides were recovered by denaturation in 0.1 M NaOH for 15 min and subsequent neutralization.

4.3 Polymerase Assembly Multiplexing (PAM) reactions

PAMs were carried out in 25 µl reactions containing 2 µl of oligonucleotide mixtures, 0.4 µM of each of the gene-end primer pairs, 1x dNTP mixture, and 0.5 µl of Advantage 2 polymerase mixture in 1x buffer (Clontech Advantage 2 PCR kit). Samples were first denatured at 95 °C for 3 min, then followed by 40-45 thermal cycles of 95 °C for 30 sec, 49 °C for 1 min, and 68 °C for 1 min kb^{-1}, and finished at 68 °C for 10 min.

ACKNOWLEDGEMENTS

We would like to thank DARPA BioComp program for early grant support of this project and DOE GTL for recent support. J.T. was supported by a LSRF fellowship.

REFERENCES

[1] Tian J, Gong H, Sheng N, Zhou X, Gulari E, Gao X, & Church GM (2004) Accurate Multiplex Gene Synthesis from Programmable DNA Chips. *Nature* 432: 1050-1054.

[2] Cello, J., Paul, A. V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science 297, 1016–1018 (2002).

[3] Smith, H. O., Hutchison, C. A. III, Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: FX174 bacteriophage from synthetic oligonucleotides. Proc. Natl Acad. Sci. USA 100, 15440–15445 (2003).

[4] 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 164, 49–53 (1995).

[5] Smith, J. & Modrich, P. Removal of polymerase-produced mutant sequences from PCR products. Proc. Natl Acad. Sci. USA 94, 6847–6850 (1997).

[6] Carr, P. A., Park, J. S., Lee, Y.-J., Yu, T., Zhang, S. & Jacobson, J. M. Protein-mediated error correction for de novo DNA synthesis. Nucleic Acids Res. (in the press).

[7] Shendure, J., Mitra, R., Varma, C. & Church, G. M. Advanced sequencing technologies: methods and goals. Nature Rev. Genet. 5, 335–344 (2004).