Published online October 11, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 18 ***5409–5417***
doi:10.1093/nar/gkh879

Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences

Xiaochuan Zhou[3,4], Shiying Cai[3], Ailing Hong[3,4], Qimin You[3], Peilin Yu[1], Nijing Sheng[1], Onnop Srivannavit[2], Seema Muranjan[3], Jean Marie Rouillard[2], Yongmei Xia[2], Xiaolin Zhang[3,4], Qin Xiang[3], Renuka Ganesh[1,4], Qi Zhu[1], Anna Matejko[1], Erdogan Gulari[2] and Xiaolian Gao[1,*]

[1]Department of Chemistry, University of Houston, Houston, TX 77004-5003, USA, [2]Department of Chemical Engineering, University of Michigan, Ann Arbor, MI 48109, USA, [3]Xeotron Co. 8275 El Rio, Suite 130, Houston, TX 77054, USA and [4]Atactic Technologies Inc., 2575 W. Bellfort, Suite 270, Houston, TX 77054, USA

Received July 21, 2004; Revised September 10, 2004; Accepted September 20, 2004

ABSTRACT

Large DNA constructs of arbitrary sequences can currently be assembled with relative ease by joining short synthetic oligodeoxynucleotides (oligonucleotides). The ability to mass produce these synthetic genes readily will have a significant impact on research in biology and medicine. Presently, high-throughput gene synthesis is unlikely, due to the limits of oligonucleotide synthesis. We describe a microfluidic PicoArray method for the simultaneous synthesis and purification of oligonucleotides that are designed for multiplex gene synthesis. Given the demand for highly pure oligonucleotides in gene synthesis processes, we used a model to improve key reaction steps in DNA synthesis. The oligonucleotides obtained were successfully used in ligation under thermal cycling conditions to generate DNA constructs of several hundreds of base pairs. Protein expression using the gene thus synthesized was demonstrated. We used a DNA assembly strategy, i.e. ligation followed by fusion PCR, and achieved effective assembling of up to 10 kb DNA constructs. These results illustrate the potential of microfluidics-based ultra-fast oligonucleotide parallel synthesis as an enabling tool for modern synthetic biology applications, such as the construction of genome-scale molecular clones and cell-free large scale protein expression.

INTRODUCTION

Modern synthetic biology is centered on the creation of new proteins and biomolecular machineries. To this end, hundreds and even thousands of genes are to be assembled and further modified to elicit and utilize their functions. As such, synthesis of long DNA of designed sequences such as minimal genomes of designated functions *in vitro* becomes crucially important. Presently, synthetic genes of designed sequences are assembled one at a time by joining the corresponding oligodeoxynucleotides (oligonucleotides) using either the polymerase chain reaction (PCR) or the ligase chain reaction (LCR) approach (1–8). The construction of a whole genome would require simultaneous synthesis of multiple genes, and thus necessitate the availability of a large number of oligonucleotides in the form of a mixture. This requirement challenges the state-of-the-art oligonucleotide synthesis, which, developed for applications that requires individual oligonucleotides, is accomplished essentially on a one-by-one basis (9). High-throughput synthesis has been implemented for simultaneous synthesis of 96–384 individual sequences (10,11). These oligonucleotides are then pooled as required for sequence assembly. The overall process, however, is time consuming and costly. Therefore for new applications, such as multiplex gene synthesis, the current methods of oligonucleotide synthesis prove the limiting step, both in terms of speed and cost. High-throughput gene synthesis is currently technologically out of reach. Additionally, the process of gene synthesis from oligonucleotides by LCR or PCR, involves a final cloning and single colony sequencing step and thus requires a much smaller amount of oligonucleotides than most conventional experiments. These requirements, of a diverse oligonucleotide pool in small amounts, for gene synthesis and potentially for other multiplex assays using oligonucleotides, call for further miniaturization of oligonucleotide synthesis in a multiplexing format. In this paper, we describe a programmable PicoArray synthesis method that is used to simultaneously synthesize the oligonucleotides designed for multiplex gene synthesis. After the synthesis on the microfluidic device, the oligonucleotides were recovered in solution and purified as a mixture. This pool of oligonucleotides was then successfully used to assemble multiple genes. This technological advancement is herein demonstrated to have the potential to overcome the

*To whom correspondence should be addressed Tel: +1 713 743 2805; Fax: +1 713 742 2709; Email: gao@mail.uh.edu

bottleneck of large scale applications of oligonucleotides such as the construction of the genes for whole genome assembly.

MATERIALS AND METHODS

PicoArray reactor

The microfluidic PicoArray reactor, which is shown in Figure 1A in a model schematic drawing, was fabricated at the Solid State Electronics Laboratory at the University of Michigan. The device was made of silicon substrate [Si(100), Wafer World] that is anodically bonded to a Corning 7740 glass wafer. The silicon layer contained three topological features: pico-reaction chambers, fluid microchannels, and inlet/outlet through-holes (Figure 1A). The actual device contained 128 × 31 (total 3698) individual pico-reaction chambers. Each three-dimensional reaction chamber measured 90 μm wide, 200 μm long, and 15 μm deep to give an internal volume of 270 pl and a reactive interior surface area of 0.045 mm^2. The overall measurements of the device were 1.7 × 2.0 cm^2 in size and 10 μl in volume. The three-dimensional features on silicon were generated using standard microelectronic fabrication procedures (12). In a typical process, a layer of photoresist (PR 1813, Hoechst Celanese) was spin-coated on a silicon wafer that was covered with a 0.6 μm thermal oxide layer, which was used as a reactive ion etching mask. The pico-reaction chamber pattern was transferred onto the surface by photolithography and buffered HF (hydrofluoric acid) etch of the thermal oxide layer, and the photoresist was removed by a resist stripper (PRS 2000, JT Baker Inc.). A new layer of photoresist (PR 1827, Hoechst Celanese) was applied; the microchannel pattern was then transferred to photoresist, and thermal oxide by photolithography and reactive ion etching. The wafer was then loaded into the deep reactive ion etching system (Surface Technology Systems, STS) to obtain ∼135 μm deep microchannels. The photoresist was removed by the resist stripper to reveal the underlying patterned layer of reaction chambers and oxide. The wafer was re-etched using the STS machine to obtain ∼15 μm deep reaction chambers and ∼150 μm microchannels. A new layer of photoresist (AZ9260, Clariant) was applied to the backside of the wafer, the inlet/outlet hole pattern was transferred to the photoresist layer by photolithography and the pin holes were etched on the STS machine. The photoresist was then stripped, the wafer was cleaned using RCA (rolling circle amplification) treatment (12), and a 0.2 μm thermal silicon oxide layer was grown on the surface to allow subsequent functionalization of the surface for chemical reactions. Finally, the microfabricated silicon wafer was anodically bonded with a glass wafer at 400°C and 1000 V (EV500 Wafer Bonder, EV Group). The fluid channels were of a tapered shape that was derived from a fluid mechanical model (CFD-ACE from CFD Research Corporation) to produce a uniform flow rate across all reaction chambers.

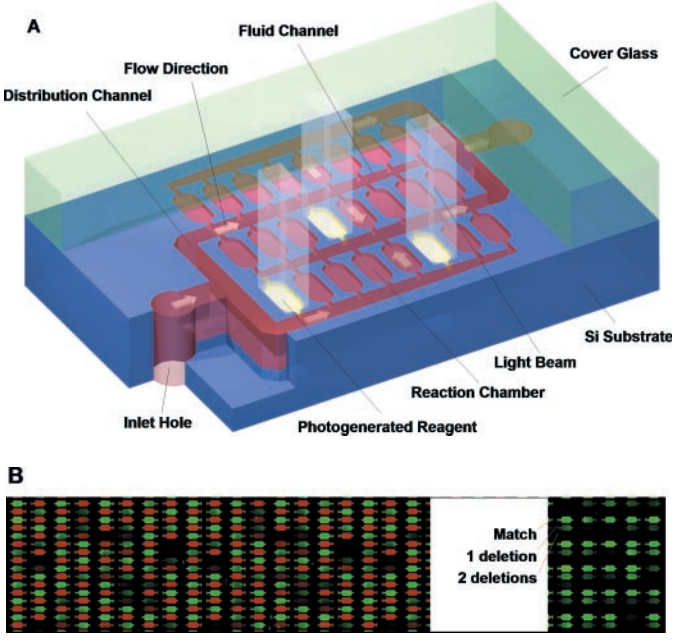
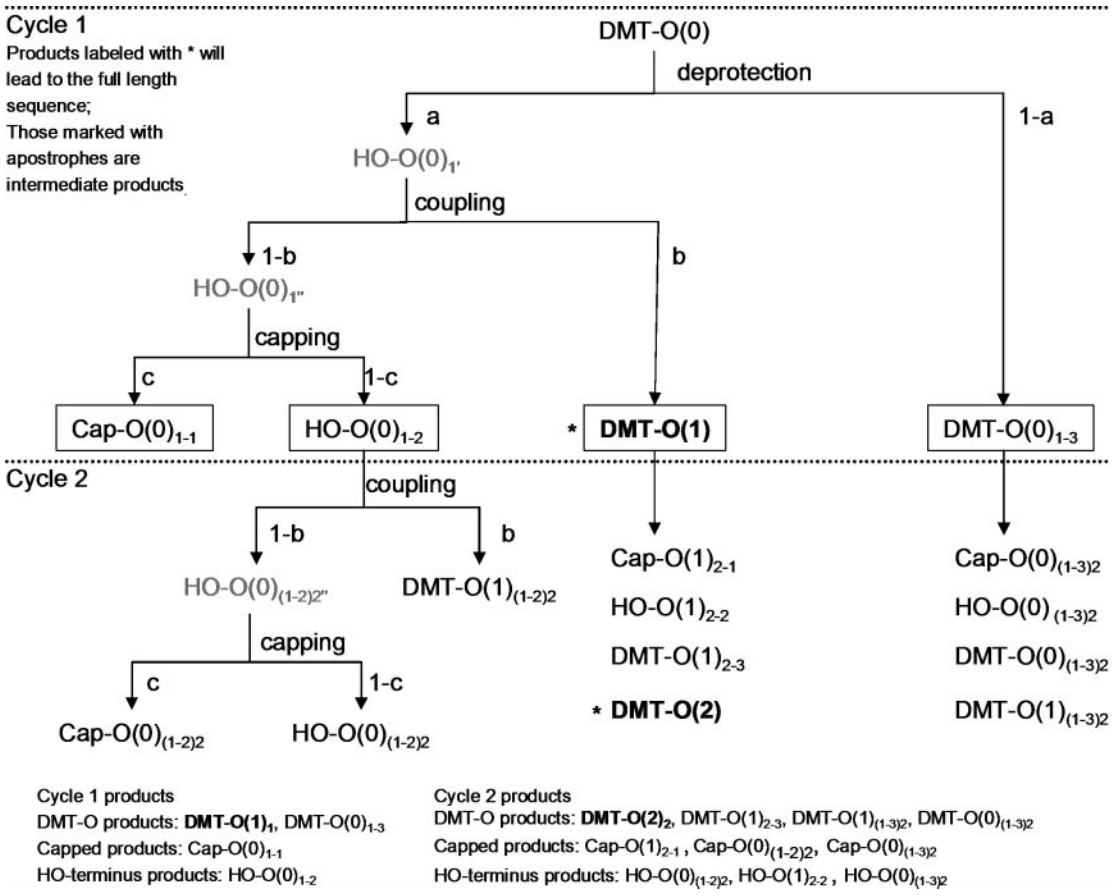


Figure 1. (**A**) Schematic illustration of a PicoArray reactor for parallel synthesis of oligonucleotides. A two-layered structure consisting of annealed silicon and glass, isolated reaction chambers etched on silicon and aligned in parallel, and inlet and outlet solution distribution channels that are connected through reaction chambers. The digital light projection is shown at selected sites to allow the PGA-controlled reaction to occur only in light-irradiated reaction chambers. (**B**) Pseudo-color image of the hybridization of the oligonucleotides that were PCR amplified from the PicoArray synthesized oligos to a detection chip. The 1011 detection probes captured either sense (red) or antisense (green) strands in a chessboard pattern. The sequences were 21–22mer, corresponding to the variable region of the target oligonucleotides. All spots have intensities above the background with spot-to-spot CV being <10%. The binding specificities are shown by the clearly defined red and green chessboard spots and the differential binding of the perfect match versus the deletion sequences (spots in the green image on right). These comparisons should be made for each set of three spots of the same column as labeled, which are of the same sequence other than the one or two base deletion sites contained in the deletion sequences. An example of the sequences of comparison sequences include: d(AACACATTAGACGGCCTCCTGC), d(AACACATTA_ACGGCCTCCTGC) with one deletion represented by '_', and d(AACACAT_AGACG_CCTCCTGC) with two deletions.

Modeling of oligonucleotide synthesis

The computational model for the calculation of product distribution of oligonucleotide synthesis considers consecutive reaction cycles consisting of reaction steps of deprotection (efficiency = a), coupling (efficiency = b), and capping (efficiency = c) (Scheme 1). In the first cycle (cycle 1) chain length $n = 0$, and the products formed from the deprotection reaction are deprotected $\text{HO-O(0)}_{1'} = a \times \text{DMT-O(0)}$ and undeprotected or unreacted $\text{DMT-O(0)}_{1-3} = (1 - a) \times \text{DMT-O(0)}$; the products from the coupling reaction are n = 1mer, coupled product $\text{DMT-O(1)} = b \times \text{HO-O(0)}_{1'}$ and uncoupled or unreacted $\text{HO-O(0)}_{1''} = (1 - b) \times \text{HO-O(0)}_{1'}$; the products from the capping reaction are capped $\text{Cap-O(0)}_{1-1} = c \times \text{HO-O(0)}_{1''}$ and uncapped or unreacted $\text{HO-O(0)}_{1-2} = (1 - c) \times \text{HO-O(0)}_{1''}$. At the end of the first cycle, the products included correct sequence DMT-O(1), capped sequence Cap-O(0)_{1-1}, HO-containing impurity sequence HO-O(0)_{1-2}, and DMT-containing impurity sequence DMT-O(0)_{1-3}. As shown in Scheme 1, the general patterns are: each DMT-O species generates four products and each HO-O species generates three products similar to that shown for cycle 1. At the end of cycle 2, there exist, in addition



Scheme 1. Analysis of the product distribution of oligonucleotide systems.

to the full-length sequence DMT-O(2), three shorter DMT-O, HO-O and Cap-O sequences. Following this analysis, at the end of the *n*th cycle of oligonucleotide synthesis, the product distribution can be expressed in the four types of sequences and their relative amounts as a function of reaction efficiencies *a*, *b*, and *c* and the number of reaction cycles (Scheme 1). The modeling program was written in Visual Basic C and it was used in Excel (Microsoft) to generate bar graphs. The representative plots of the model are shown in Figure 2.

PicoArray oligonucleotide synthesis

The oligonucleotides used for ligation-based gene synthesis were derived from the sequence of the desired DNA construct using Gene2Oligo (13) (http://berry.engin.umich.edu/gene2oligo/), DNAWorks (14), or a macro-script of Excel (Microsoft) written in-house. An example of the sequence design is given in Supplementary Materials. The major considerations in sequence design include length, T_m(melting temperature), cross-interactions, and secondary structure formation, given the restriction of the gene sequences to be synthesized. The oligonucleotides were on an average, 30 or 45mer fragments with cohesive joints (Figure 4D) (13). A pseudo-duplex forms and the top strand has nicks at positions 15, 45, etc. and the bottom strand has nicks at positions 30, 60, and so on. The T_m of oligonucleotides, without taking into account the presence of a nick in the middle of the duplex, was adjusted to comparable values by relaxing the requirement of the length. The DNA constructs assembled in this work include enhanced green fluorescent protein (EGFP) (gi:603191, 742 bp), enhanced yellow fluorescent protein (EYFP) (gi:7638256, 712 nt), DNA gyrase subunitB (BB0436), LacZ fragment in the cloning vector (512 nt), and an ensemble of these constructs containing silent mutations.

The PicoArray reaction sites (the internal surfaces of the reaction chambers) were first functionalized with 4′,4-dimethoxytrityl (DMT) groups by coupling to 5′-DMT-thymidine and the PicoArray was enclosed in a holder that was connected to an Expedite 8909 DNA synthesizer via luer-lock joints. The syntheses used standard phosphoramidite chemistry (15) except for the deprotection step, where photogenerated acid (PGA) was used to deprotect the DMT group at selected reaction sites (16,17). The synthesizer is equipped with a programmable optical unit that uses digital photolithographic projection via a Digital Light Projector (DLP^{TM}, Texas Instruments) (18). At each deprotection step, the PicoArray was filled with the PGA precursor solution (in CH_2Cl_2); a predetermined light pattern was automatically projected onto the device surface to trigger the formation of PGA. The wavelength used was 405 nm, the light intensity on the device surface was 7 mW/cm^2, and the exposure time was 2–3 s. The synthesis protocol was similar to what was used for regular DNA oligonucleotide synthesis but is optimized with regard to the microfluidic flow rate and volume, to

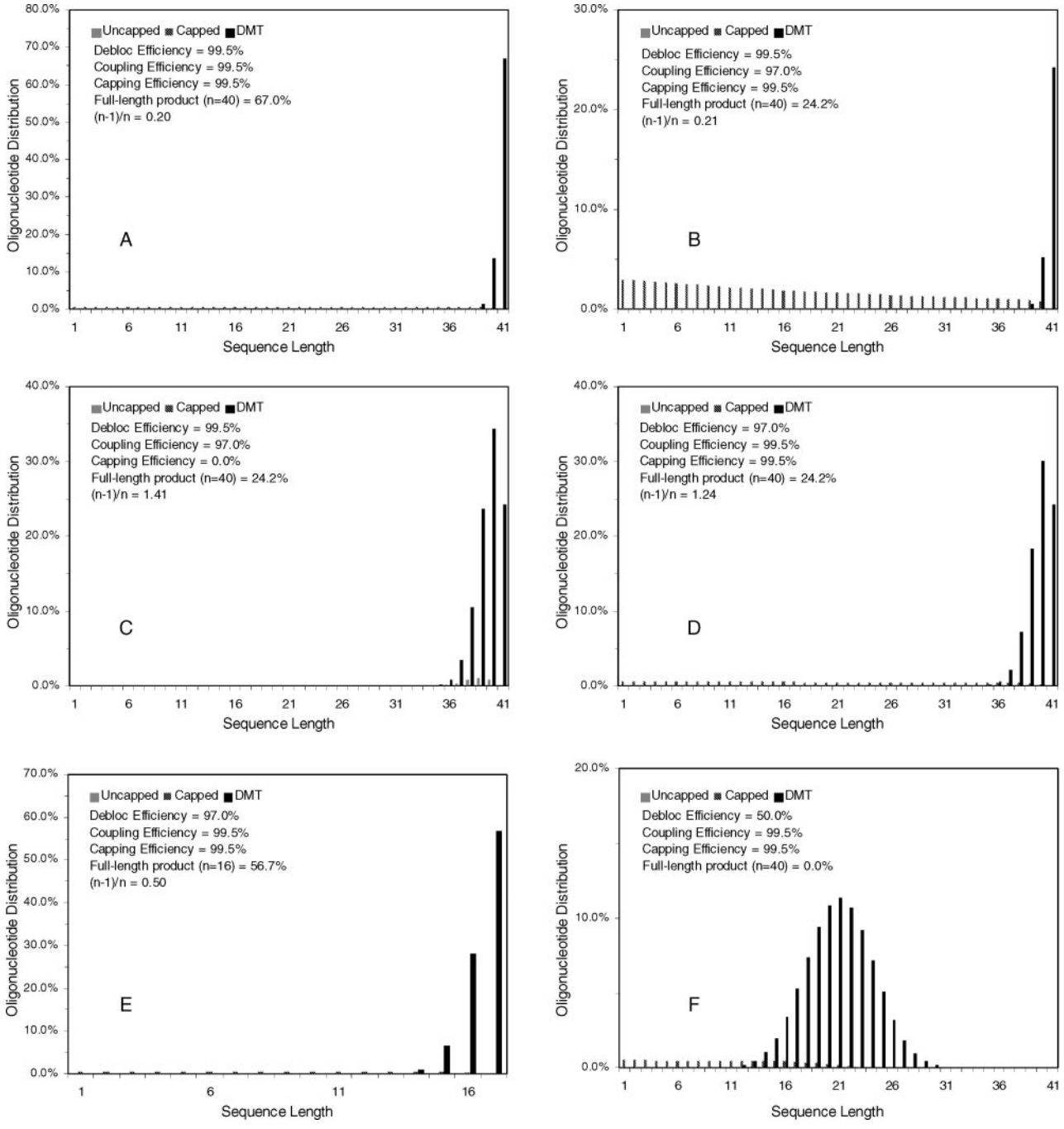


Figure 2. The predicted product distributions of oligodeoxynucleotide synthesis by modeling. The length of the oligomers is given as the *x*-axis. (**A**) In a case of high yield synthesis, the full-length 40mer sequence is 67.0% (40-cycle synthesis) and the $(n-1)$mer (39mer accumulated from the various failure steps) over *n*mer (the full-length sequence) is 0.20. (**B**) A reaction with a lower coupling efficiency would result in a low yield of full-length product (24.2%) as compared to 67.0% in (A), but this would only marginally increase $n-1$:*n* ratio of sequences from 0.20 to 0.21. The majority of the failure products are capped as short fragments. (**C**) The same reaction efficiencies as in (B) but without using capping. This results in a poor product distribution; $n-1$:*n* ratio is 1.41, which is significantly higher than 0.21 shown in (B) for synthesis using the capping step. (**D**) A reaction with a low deprotection (or deblock) efficiency of 97.0% but high coupling and capping efficiencies would still produce a poor product distribution having a high $n-1$:*n* ratio of 1.24. (**E**) The same reaction as in (D) but for a 16mer synthesis. The full-length sequence is 24.2%. (**F**) In a dramatic case of poor deprotection, no full-length sequence can be produced and the impurity products show characteristic binominal distribution patterns.

ensure high efficiency reactions. When needed for ligation, 5′-phosphate was added by final coupling with the 5′-phosphorylating agent (Glen Research) to the designated oligonucleotides; for fluorescent dye labeling, the last step of coupling used fluorescein phosphoramidite or other dye molecules according to the procedures recommended by the vendor (Glen Research). The fluorescent images of the directly labeled oligonucleotides were acquired on an Axon laser scanner (GenePix 4000B); signal intensities were obtained using ArrayPro (Media Cybernetics).

Recovery and CE analysis of the oligonucleotides obtained from the PicoArray Reactor

The deprotection of nucleobases was accomplished using a solution of 0.5 ml EDA/EtOH (50/50 by volume) that circulated through the reactor for 2 h followed by EtOH wash. Byproducts were removed by circulation with washing solutions [EtOH, 0.5 M sodium acetate, and 6× SSPE (saline-sodium phosphate-EDTA) buffer]. The oligonucleotide cleavage reaction used ~80 µl of aqueous ammonia solution at 37°C for 2 h. The solution was eluted from the reactor in ~100 µl and collected in a microcentrifuge tube. The elute was concentrated to ~20 µl using microcon YM3 spin-column (Millipore). The completion of the cleavage in some test cases was monitored by direct fluorescent dye labeling, with images taken before and after the cleavage. The cleavage solution was mixed with an equal volume of phenol:chloroform:isoamyl alcohol (Sigma-Aldrich) and the aqueous layer was collected; this extraction was repeated twice. The combined aqueous layer solutions were washed with ethyl ether and the volume reduced to ~20 µl. Oligonucleotides were precipitated with NaOAc/EtOH, followed by standing overnight at −70°C, and collected as pellets after centrifugation. For capillary electrophoresis (CE) quality analysis, a single sequence [31mer, d(GGCGATTAAGTTGGGTCGTCTGAGCATCTGA)] was synthesized using PGA on PicoArray and the sequence was deprotected and cleaved using aqueous ammonia. The sample was analyzed according to vendor recommended procedures (P/AGE MDQ CE system, Beckman Coulter). The CE chromatograph is shown in Figure 3.

Oligonucleotide point check by hybridization

All solutions used were filtered through a 0.45 µm filter. The detection chip that contained the complementary sequences of the oligonucleotides were used for gene assembling. A micro-peristaltic pump was used to circulate the prehybridization solution (6× SSPE, 25% formamide and 0.2% BSA). The oligonucleotides obtained from the microfluidic PicoArray synthesis were dissolved in ~100 µl hybridization solution and were circulated for ~8 h or more at a flow rate of 0.5 ml/min at 22°C for 30mer oligonucleotides, or at 32°C for longer sequences. The images of hybridization were acquired using an Axon Genpix 400B scanner equipped with 532 and 635 nm excitation lasers and processed by using ArrayPro. We did not observe the cy5 signal decay for the image scans acquired during the hybridization and washes as reported (19). It is believed that the ozone is mainly responsible for the destruction of the dye. Since the microfluidic chip is a closed system the hybridized dye-containing DNA samples would essentially have no exposure to atmospheric ozone.

Ligation assembly of DNA constructs and EGFP expression

For a 10 µl reaction, 5 µl of the oligonucleotide solution was used with 1 µl of 10× Taq ligation buffer and 0.5 µl of *Taq* ligase (New England BioLabs) (20 U). The mixture was heated to 95°C for 2 min, and followed by 40 cycles at 75°C for 1 min and 45–60°C for 5 min for melting and annealing/ligation, respectively. The reaction was terminated by lowering the temperature to 4°C. The ligation products were divided into several portions and each was PCR-amplified with a high-fidelity polymerase (*Pfu*Ultra, Stratagene) using several primer pairs specific for amplifying different regions of the ligated sequence. The PCR products were analyzed by 1.5% agarose gel (LE agarose, analytical grade, Promega) (Figure 4A). The DNA band of desired size was recovered from the gel for cloning and sequencing. For sequencing, the ligated EGFP sequence was inserted into pCR4-TOPO (Invitrogen) and transformed into *E.coli* One Shot® TOP10 competent cells (Invitrogen). For gene expression, the synthesized and amplified DNA sequence was cloned into pTrcHis vector and transformed into *E.coli* XL1-blue competent cells. The transformants were transferred to grid plates and induced by IPTG (isopropyl-beta-D-thiogalactopyranoside). The fluorescent colonies were observed under UV (ultraviolet) light. Separately, the EGFP gene (Clontech) was cloned into pTrcHis vector as a positive control.

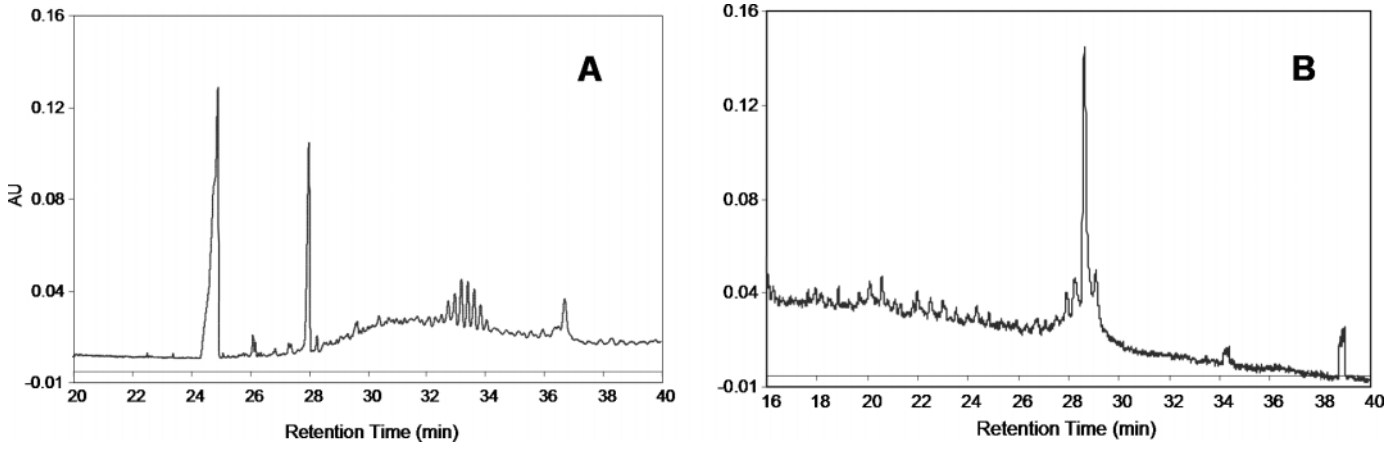


Figure 3. Plots of CE of oligonucleotides synthesized on a PicoArray reactor using PGA. (**A**) CE plot of a 31mer oligonucleotide synthesized as a case of poor deprotection efficiency (see DNA synthesis modeling, Figure 2F). (**B**) CE plot of a 31mer oligonucleotide synthesized after reaction optimization and high yield synthesis (98.8%) was achieved using the microfluidic PicoArray reactor.

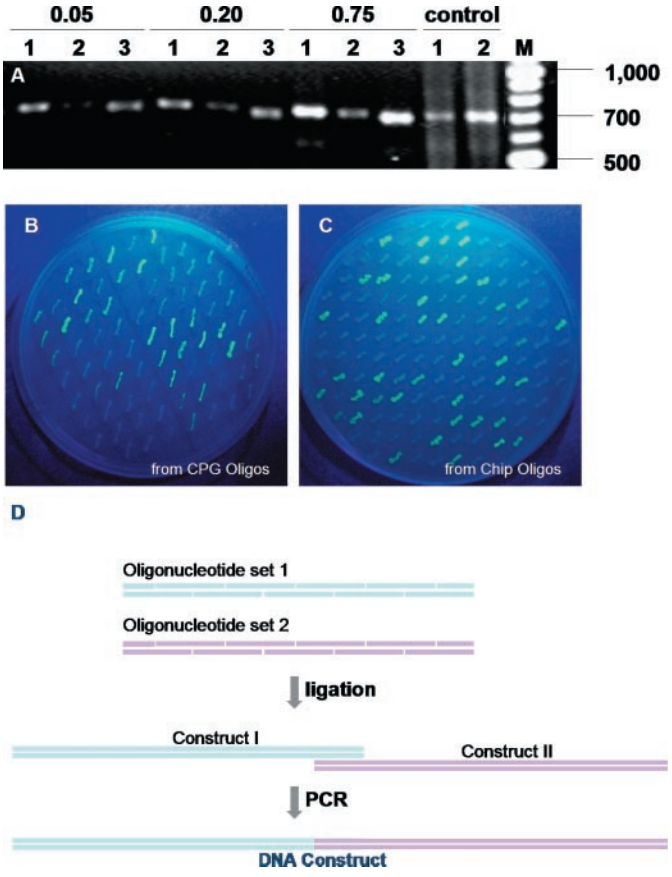


Figure 4. (**A**) Electrophoresis gel of EGFP gene (714 bp) synthesized using ligation, followed by PCR amplification using primers for the 714 bp region. Labels 0.05, 0.20, and 0.75 are fractions of the oligonucleotides used for ligation reactions based on the material collected from the PicoArray synthesis. In each set, lane 1 used Pfu DNA polymerase, lane 2 used Taq (SureStart), lane 3 used a different set of primers from lane 1 and Pfu. Control lanes used oligonucleotides synthesized on CPG. M represents the 1 kb molecular marker. (**B** and **C**) Displays of the EGFP protein expressed from genes assembled using (B) oligonucleotides (synthesized on CPG) from a commercial source or (C) oligonucleotides synthesized and recovered from the PicoArray reactor. The number of the functional colonies in (A) is 26.8% (19 out of 71) and in (B) is 30.0% (40 out of 132), excluding two positive and two negative controls. (**D**) Illustration of the overall DNA synthesis strategy, i.e. ligation of oligonucleotides into fragments of several hundreds of base pairs followed by fusion PCR, used in this work.

Quantitation (real-time PCR) measurement of oligonucleotides synthesized on the PicoArray reactor

A quantitative real-time PCR of the oligonucleotides obtained from PicoArray synthesis was performed on an ABI Prism 7000 Sequence Detection System in the presence of SYBR-Green fluorescent dye. The PCR reaction was 25 µl per well with each performed in triplicate in a 96-well plate by following the manufacturer's instructions (Applied Biosystems). Typically, 50 ng/ml of yeast tRNA was present in all reactions; oligonucleotides from PicoArray synthesis were diluted by 10 to 10 000-fold to desired concentrations; a known amount of a 45mer sequence was also diluted to a series of concentrations and used as a reference to normalize cycle threshold (Ct). DNA dilution buffer was used as the no template control. Two microliters of the diluted oligonucleotides were added to wells that contained the premixed solution, consisting of specific primers, detector and enzyme (Sure Start® *Taq* polymerase and Brilliant® SYBR® Green QPCR Master Mix, Strategene). The reaction cycles used were 50°C for 2 min, 95°C for 10 min, and then 40 cycles at 95°C for 15 s and 55°C for 1 min. In some cases, reaction products were further analyzed by CE gel electrophoresis and slab gel electrophoresis. The number of oligonucleotides per reaction well was calculated based on Ct, and from the calibration of the reference oligonucleotide.

RESULTS

The Microfluidic PicoArray Synthesis

In the programmable PicoArray synthesis method, a new microfluidic device is used as a miniaturized synthesizer for solid support parallel synthesis of oligonucleotides. This device is a multiplexing reactor that is configured with massively parallel picoliter-scale reaction chambers as illustrated in Figure 1A. The microfluidic synthesis device is programmable and is readily adopted by the existing automated synthesizers. During the operation, fluid, driven by pressure, flows into the device through an entrance through-hole, splits evenly into the entry binary distribution channels, flows down the tapered fluid channels, passes through the reaction chambers in parallel, enters the tapered fluid channels on the other side of the reaction chambers, emerges in the exit binary distribution channels, and exits through an exit through-hole. The entirely sealed, positive pressure operational system helps to isolate the cross-interactions among the reaction chambers, such as *in situ* generated acid diffusion (see next paragraph), and avoid trapping of gas bubbles, solvent evaporation, air-borne contamination, and air-oxidation. The synthesis device is compatible with organic solvents used for oligonucleotide synthesis.

The PicoArray is designed to enhance the performance of a novel PGA chemistry (16), which was demonstrated for *in situ* synthesis of DNA microarray chips on non-wetting film patterned glass plates (17,20). The synthesis process was carried out on a regular DNA synthesizer that is equipped with a programmable digital light projector as described previously (17,20,21). The PicoArray microfluidic device was connected to the synthesizer in the same way as a regular synthesis column and the synthesis was similar to the process that was described previously (17). In a typical experiment, oligonucleotide synthesis started on the silane linker derivatized surfaces that contained DMT protected hydroxyl. The reagent flow rate and the synthesis protocols were optimized for high yield and minimal reaction time. Multiple oligonucleotides are synthesized in parallel by gating the reaction using PGA, which was generated upon light irradiation at selected reaction chambers, to remove the DMT group at those sites, thereby allowing nucleophosphoramidite coupling (chain growth) in each reaction cycle. The PicoArray microfluidic device exhibits excellent mass transfer properties. For instance, complete deprotection of the acid labile group from the 5′-terminus of the growing chain took 2–3 s, compared with the 60 s that was required for a confined droplet on a glass plate in our previous studies (17). Under the current reaction conditions, the reaction cross-talk of the adjacent cells is minimal and the composition integrity (sequence fidelity) inside an individual reaction chamber is maintained.

The PicoArray synthesis device has only 10 μl flow-through volume, and thus the chemical consumption for synthesizing an array of oligonucleotides is comparable with regular DNA oligonucleotide synthesis of a single sequence. This translates into micro-amounts of reagents and solvents on a per sequence basis as thousands of oligonucleotides are simultaneously made in a PicoArray synthesis device.

Modeling of the synthesis

To better understand the results and enhance the quality for the PicoArray synthesis, a computational model was created to simulate the product distribution of oligonucleotides as a function of the efficiencies of three individual reaction steps: deprotection using acid, coupling using nucleophosphoramidite, and capping using acetic anhydride (Scheme 1). The synthesis of oligonucleotides consists of repeated cycles, with each cycle performing the three individual reaction steps plus the oxidation reaction. The products under consideration by the model include the full-length sequence, the sequences terminated with OH (uncapped), the sequence terminated with the protecting group (DMT) and the capped sequences, [Cap-O($n-1$), Cap-O($n-2$), etc.], where each category of the sequence is a mixture of different lengths accumulated over multiple reaction cycles. This model reveals distinct patterns of product distribution as a function of the efficiency of individual reaction steps (Figure 2). These results are used to isolate the factors that most affect the quality of synthesis, especially those that influence the formation of the *n*-1mer oligonucleotides. This is particularly important because of the difficulty in the discrimination between the *n*-1mer and the *n*mer oligonucleotides. This is most probably the primary source of errors in the subsequent experiments. The model predicted that besides high yield, stepwise synthesis is critical to the final amount of full-length sequences (Figure 2A versus 2B), but high efficiency deprotection is most critical in reducing *n*-1 and other undesirable impurity sequences, the presence of which significantly reduces the quality of the synthesis. Low efficiency deprotection would result in shifting of the product distribution to shorter sequences. For example, a reaction condition having coupling and capping efficiencies of 99.5% and a moderate deprotection efficiency of 97.0% would result in (*n*-1) versus *n* ratios of 0.50 and 1.24 for the synthesis of 16 and 40mer sequences, respectively, rendering the process unsuitable for making 40mer sequences (Figure 2E and D). Experimental data shown in Figure 3A validated the model (Figure 2F) when poor deprotection efficiency was created in a synthesis. On the other hand, lower coupling efficiency may be compensated by efficient capping; such synthesis results in fewer full-length sequences (24.2% compared with 67.0% of an ideal synthesis) but without a significant increase of the n-1 versus *n*mer ratio (a ratio of 0.21 compared to 0.20 of an ideal synthesis) (Figure 2B and A). These comparisons clearly illustrate the importance of efficient capping for reducing the formation of *n*-1mer sequences (Figure 2C). Using the modeling and product profile analysis as a guide, we optimized synthesis conditions in the PicoArray reaction device and routinely achieved 98.8% or better stepwise yield (Figure 3B). The synthesis efficiency would generate 58% full-length 45mer or 70% full-length 30mer sequences. Oligonucleotides >100 nt long (30% full length by prediction) were synthesized using the PicoArray device, cloned and verified by sequencing to contain the correct sequences (data not shown).

Post-synthesis analysis and reverse hybridization of oligonucleotides

The synthesis results were analyzed using several methods including the use of CE as described above. Using real-time PCR, we selectively measured several sequences in the cleaved oligonucleotide mixture and derived an average quantity of 3×10^9 (or 5 fmol) sequences per reaction chamber. This suggests that a total of 20 pmol full-length sequences were recovered from each PicoArray synthesis. The overall presence of the synthesized oligonucleotides in the cleaved mixture product was verified by hybridizing the oligonucleotide mixture to a detection DNA array chip that contained complementary sequences (i.e. oligonucleotide point check) for which the probes have comparable duplex melting temperature (T_m) (13). In the experiment shown in Figure 1B we amplified the cleaved oligonucleotide mixture using PCR (the sequences contain priming sites on both ends), and hybridized the PCR product of both sense (labeled with Cy5 and shown in red) and antisense (labeled with Cy3 and shown in green) strands to the detection DNA array chip that contained 1011 probes in duplicate. In a separate experiment, the target duplexes were products of PCR and T7 *in vitro* transcription of the oligonucleotides synthesized and recovered from the PicoArray reactor. The presence of complete sets of both strands was confirmed by the observation of an expected red and green chessboard pattern that alternated between hybridized sense and antisense sequences. Nearly 100% of the expected hybridization sites of the cleaved oligonucleotides have intensities above the average background plus three times of the standard deviation of the background; the integrity of the individual sequences obtained from PCR was assessed by hybridization as well. The complementary sequences (perfect match) were easily distinguished from their counterparts that contained mismatched and deleted bases, by their higher intensities. In separate experiments, the hybridized oligonucleotides can be recovered from the surface under denaturing conditions and used in ligation and PCR reactions. These results demonstrate the potential of massively parallel purification of T_m equalized oligonucleotides by hybridization selection of high-fidelity sequences under optimized experimental conditions. In a separate paper, Church and co-workers provide further experimental details on oligonucleotide point check by FAM labeling and the use of oligonucleotides in PCR and multiplex gene synthesis; they demonstrate that hybridization purification that used specific probes, synthesized on the PicoArray reactor efficiently reduces the error rate from a level of ~1/160 to 1/1394 (G. Church, J. Tian, H. Gong, N. Sheng, X. Gao, X. Zhou and E. Gulari, manuscript submitted).

Assembling of oligonucleotides

The general strategy of DNA synthesis used in this work is depicted in Figure 4D,which involves ligation of oligonucleotides containing cohesive ends, followed by fusion PCR. Multiple DNA constructs (200–2000 total bp) were assembled from oligonucleotide mixtures of 30 or 45mers that were

obtained using the PicoArray synthesis method. An example of a 1 kb gene sequence and its oligonucleotides is given in the Supplementary Material. The PicoArray oligonucleotides were collected in a 20 µl volume, and the materials produced from one synthesis were sufficient for 2–4 ligation reactions. The ligation products were then PCR amplified and visualized by agarose gel electrophoresis (Figure 4A). The longest DNA constructs directly assembled by ligation were 714 bp EGFP and 712 bp EYFP DNA fragments. A 1040 bp EGFP gene (Supplementary Material) plus flanking sequences was assembled from overlapping 480 and 580 bp DNA ligation fragments, assembled simultaneously, followed by fusion PCR. Simultaneous assembling of five DNA constructs of a total of 1.2 kb from 80 oligonucleotides from the same PicoArray synthesis was also demonstrated; the longest overall length of these DNA fragments was 2 kb. To explore the potential of large DNA synthesis by the ligation–fusion PCR strategy, experiments were performed using ~700 bp fragments (total 15, 5 from LacZ, 4 from luciferase, 2 from tubulin, 1 from BGH, 1 from Amp, 1 from YFP and 1 from RsRed) to assemble 2 kb DNA constructs and then to assemble a 10 kb DNA construct, which was validated by gel electrophoresis (data not shown). The known difficult gene, DNA gyrase subunit B (1.9 kb, 67.8% AT), was synthesized by ligation and fusion PCR. The discussed genes synthesized were validated by conventional sequencing. To avoid cross contamination associated with the PCR reactions used in DNA synthesis, silent mutations of the genetic code were incorporated into the various positions of the various DNA sequences to distinguish them from among the different synthesis experiments. Figure 4B shows the expression of EGFP from synthetic genes from the commercial or the PicoArray synthesized oligonucleotides, and 26.8 versus 30.0% of colonies contained functional full-length genes and expressed EGFP, respectively. The sequencing results gave an error rate of 1.6‰ (‰ is per thousand) for the PicoArray synthesized EGFP gene that was derived, which is comparable to the 1.7/‰ error rate of the synthetic gene that was assembled from commercial oligonucleotides. On an average, the error rate of the synthetic genes based on PicoArray oligonucleotides was found to be between 1 and 5/‰ when no post-synthesis purification was applied. The types of errors were in most cases deletion > substitution > insertion, and the distribution of these errors was somewhat random. In addition, synthesis efficiency is another source of error; further understanding of the sources of sequence errors would require a large amount of sequencing data accumulated from DNA synthesis. Whereas the attained error rate is sufficient for 1 kb gene synthesis, it must be reduced by a factor of 10 for the synthesis of larger genes.

CONCLUSIONS

We have demonstrated a microfluidic PicoArray device for miniaturization of synthesis and a method that enables ultrafast generation of a large number of oligonucleotides in high quality. Our microfluidic synthesis device demonstrates several advantages: (a) it is readily adopted by the existing automated synthesizers for parallel synthesis of oligonucleotides; (b) it is programmable for synthesis of any desired sequence; (c) it is simple to operate; (d) it consumes only small amounts of reagents and solvents on a per sequence basis; (e) it displays significant kinetic acceleration compared with that of conventional reactions; (f) it allows easy recovery of sequences synthesized in a small volume for direct use in the subsequent enzymatic reactions; (g) it is a closed system that can protect the contained samples from contamination and air-oxidation. While simple hybridization on chip can tolerate the presence of high percentages of impure sequences, for multiplex experiments such as multiplex gene assembly, the quality of the synthesis is much more critical. We used a model of DNA synthesis to improve the key step of the reaction cycles; the oligonucleotides obtained were directly used for ligation under thermal cycling conditions to give multiple DNA fragments of several hundreds of base pairs. Long DNA constructs up to 10 kb were assembled by ligation followed by fusion PCR. These promising results set the stage for proteins of arbitrary sequences to be synthesized rapidly and for minimal genome synthesis to become routine. Furthermore, our programmable chemistry platform coupled with the microfluidic PicoArray provides a new avenue to meet the increasingly complex synthetic demands of large scale biology. Such examples include synthetic biology and the construction of genomic-scale siRNA vectors to accelerate the functional screening of genes based on RNA interference (RNAi) mechanisms (22). The method can also be coupled with cell-free protein expression (23,24) to greatly accelerate the synthesis of protein libraries. The PicoArray synthesis method presented herein is scalable to the parallel generation of a million sequences. In this regard, the microfluidic configuration has a unique advantage in that, a large number of oligonucleotides simultaneously synthesized can easily be divided after cleavage and collected as subsets of sequence mixtures of reasonable sizes. This is more amenable to purification and handling, as the complexity of the oligonucleotide mixtures can be controlled to ensure specificity in the subsequent multiplex reactions and enzymatic assays.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful for the support of the application and engineering groups at Xeotron Co. and for the many valuable and inspiring discussions with Professor George Church at Harvard University. This research is supported by grants from NIH and DARPA and the R.A. Welch Foundation (E-1270).

REFERENCES

1. Stemmer,W.P.C., Crameri,A., Ha,K.D., Brennan,T.M. and Heyneker,H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, **164**, 49–53.
2. Barany,F. (1991) Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc. Natl Acad. Sci. USA*, **88**, 189–193.
3. Dillon,P.J. and Rosen,C.A. (1990) A rapid method for the construction of synthetic genes using the polymerase chain reaction. *Biotechniques*, **9**, 298–300.

4. Bambot,S.B. and Russell,A.J. (1993) Efficient total gene synthesis of 1.35-kb hybrid alpha-lytic protease gene using the polymerase chain reaction. *PCR Methods Appl*., **2**, 266–271.
5. Sandhu,G.S. Aleff,R.A. and Kline,B.C. (1992) Dual asymmetric PCR: one-step construction of synthetic genes. *Biotechniques*, **12**, 14–16.
6. Barany,F. (1991) The ligase chain reaction in a PCR world. *PCR Methods Appl*., **1**, 5–16.
7. Au,L.-C., Yang,F.-Y., Yang,W.-J., Lo,S.-H. and Kao,C.-F. (1998) Gene synthesis by a LCR-based approach : High-level production of leptin-L54 using synthetic gene in *Escherichia coli*. *Biochem. Biophys. Res. Commun*., **248**, 200–203.
8. Smith,O.H., Hutchison,C.A., Pfannkoch,C. and Venter,J.C. (2003) Generating a synthetic genome by whole genome assembly: X174 bacteriophage from synthetic oligonucleotides. *Proc. Natl Acad. Sci. USA*, **100**, 15440–15445.
9. Beaucage,S.L (1993) Oligodeoxyribonucleotide synthesis: phosphoramidite approach. In Agrawal,S. (ed.), *Protocols for Oligonucleotides and Analogs*. Methods in Molecular Biology series, Vol. 20, Humana Press, New Jersey, pp. 33–62.
10. Lebl,M., Burger,C., Ellman,B., Heiner,D., Ibrahim,G., Jones,A., Nibbe,M., Thompson,J., Mudra,P., Pokorny,V., Poncar,P. and Zenisek,K. (2001) Fully automated oligonucleotide synthesizer. *Coll. Czechoslovak Chem. Commun*., **66**, 1299–1314.
11. Lashkari,D.A., Hunicke-Smith,S.P., Norgren,R.M., Davis,R.W. and Brennan,T. (1995) An automated multiplex oligonucleotide synthesizer: development of high-throughput, low-cost DNA synthesis. *Proc. Natl Acad. Sci. USA*, **92**, 7912–7915.
12. Madou,M.J., (2002) *Fundamentals of Microfabrication: The Science of Miniaturization*, 2nd edn. CRC Press, Boca Raton, FL.
13. Rouillard,J.M., Lee,W.Truan,G., Gao,X., Zhou,X. and Gulari,E. (2004) Gene2Oligo: Oligonucleotide design for *in vitro* gene synthesis. *Nucleic Acids Res*., **32**, W176–180.
14. Hoover,D.M. and Lubkowski,J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res*., **30**, e43.
15. McBride,L.J. and Caruthers,M.H. (1983) An investigation of several deoxynucleotide phosphoramidites. *Tetrahedron Lett*., **24**, 245–248.
16. Gao,X., Yu,P.Y., LeProust,E., Sonigo,L., Pellois,J.P. and Zhang,H. (1998) Oligonucleotide synthesis using solution photogenerated acids. *J. Am. Chem. Soc*., **120**, 12698–12699.
17. Gao,X., LeProust,E., Zhang,H., Srivannavit,O., Gulari,E., Yu,P., Nishiguchi,C., Xiang,Q. and Zhou,X. (2001) Flexible DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Res*., **29**, 4744–4750.
18. Hornbeck,L.J. (1996) Digital light processing and MEMS, reflecting the digital display needs of the networked society. *SPIE Eur. Proc*., **2783**, 135–145.
19. Fare,T.L., *et al*. (2003) Effects of atmospheric ozone on microarray data quality. *Anal. Chem*., **75**, 4672–4675.
20. Gao,X., Gulari,E. and Zhou,X. (2003) *In situ* synthesis of oligonucleotide microarrays. *Biopolymers*, **73**, 579–596.
21. Pellois,J.P., Zhou,X., Srivannavit,O., Zhou,T., Erdogan,G. and Gao,X. (2002) Individually addressable parallel peptide synthesis on microchips. *Nat. Biotechnol*., **20**, 922–926.
22. Paddison,P.J., Silva,J.M., Conklin,D.S., Schlabach,M., Li,M., Aruleba,S., Balija,V., O'Shaughnessy,A., Gnoj,L., Scobie,K., *et al*. (2004) A resource for large-scale RNA-interference-based screens in mammals. *Nature*, **428**, 427–431.
23. Yokoyama,S. (2003) Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol*., **7**, 39–43.
24. Andreadis,J.D. and Chrisey,L.A. (2000) Use of immobilized PCR primers to generate covalently immobilized DNAs for *in vitro* transcription/translation reactions. *Nucleic Acids Res*., **28**, e5.