

Light-generated oligonucleotide arrays for rapid DNA sequence analysis

(sequencing by hybridization/combinatorial chemistry/DNA diagnostics)

ANN CAVIANI PEASE[†], DENNIS SOLAS[†], EDWARD J. SULLIVAN[†], MAUREEN T. CRONIN[‡], CHRISTOPHER P. HOLMES[†], AND STEPHEN P. A. FODOR^{†‡}

[‡]Affymetrix, 3380 Central Expressway, Santa Clara, CA 95051; and [†]Affymax, 4001 Miranda Avenue, Palo Alto, CA 94304

Communicated by Ronald W. Davis, January 4, 1994

ABSTRACT In many areas of molecular biology there is a need to rapidly extract and analyze genetic information; however, current technologies for DNA sequence analysis are slow and labor intensive. We report here how modern photolithographic techniques can be used to facilitate sequence analysis by generating miniaturized arrays of densely packed oligonucleotide probes. These probe arrays, or DNA chips, can then be applied to parallel DNA hybridization analysis, directly yielding sequence information. In a preliminary experiment, a 1.28×1.28 cm array of 256 different octanucleotides was produced in 16 chemical reaction cycles, requiring 4 hr to complete. The hybridization pattern of fluorescently labeled oligonucleotide targets was then detected by epifluorescence microscopy. The fluorescence signals from complementary probes were 5–35 times stronger than those with single or double base-pair hybridization mismatches, demonstrating specificity in the identification of complementary sequences. This method should prove to be a powerful tool for rapid investigations in human genetics and diagnostics, pathogen detection, and DNA molecular recognition.

Conventional DNA sequencing technology is a laborious procedure requiring electrophoretic size separation of labeled DNA fragments. An alternative approach to *de novo* DNA sequencing, termed sequencing by hybridization (SBH), has been proposed (1–3). This method uses a set of short oligonucleotide probes of defined sequence to search for complementary sequences on a longer target strand of DNA. The hybridization pattern is then used to reconstruct the target DNA sequence. It is envisioned that hybridization analysis of large numbers of probes can be used to sequence long stretches of DNA. In more immediate applications of hybridization methodology, a small number of probes can be used to interrogate local DNA structure.

The strategy of SBH can be illustrated by the following example. A 12-mer target DNA sequence, AGCCTAGCTGAA, is mixed with a complete set of octanucleotide probes. If only perfect hybridization complementarity is considered, 5 of the 65,536 octamer probes—TCGGATCG, CGGATCGA, GGATCGAC, GATCGACT, and ATCGACTT—will hybridize to the target. Alignment of the overlapping sequences from the hybridizing probes reconstructs the complement of the original 12-mer target:

```
TCGGATCG
   CGGATCGA
     GGATCGAC
       GATCGACT
         ATCGACTT
           TCGGATCGACTT
```

Hybridization methodology can be carried out by attaching target DNA to a surface. The target is then interrogated with a set of oligonucleotide probes, one at a time (4, 5). This approach can be implemented with well-established methods of immobilization and hybridization detection but involves a large number of manipulations. For example, to probe a sequence utilizing a full set of octanucleotides, tens of thousands of hybridization reactions must be performed.

Alternatively, SBH can be carried out by attaching probes to a surface in an array format where the identity of the probe at each site is known. The target DNA is then added to the array of probes. The hybridization pattern determined in a single experiment directly reveals the identity of all complementary probes. The testing of this SBH format has required the development of new technologies for the fabrication of oligonucleotide arrays (6, 7). Most recently, an oligonucleotide array of 256 octanucleotides was generated using a solution-channeling device to direct the oligonucleotide probe synthesis into 3×3 mm sites (7). In this format, the resolution limit of approximately 1×1 mm may ultimately limit the number of probes that one can synthesize on a substrate of a manageable size (7).

We report here a method by which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays. Photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry, and versatile combinatorial synthesis strategies have been developed for this technology. A matrix of 256 spatially defined oligonucleotide probes was generated, and the ability to use the array to identify complementary sequences was demonstrated by hybridizing fluorescently labeled octanucleotides. The hybridization pattern demonstrates a high degree of base specificity and reveals the sequence of oligonucleotide targets.

MATERIALS AND METHODS

The basic strategy for light-directed oligonucleotide synthesis (6) is outlined in Fig. 1. The surface of a solid support modified with photolabile protecting groups (X) is illuminated through a photolithographic mask, yielding reactive hydroxyl groups in the illuminated regions. A 3'-*O*-phosphoramidite-activated deoxynucleoside (protected at the 5'-hydroxyl with a photolabile group) is then presented to the surface and coupling occurs at sites that were exposed to light. Following capping, and oxidation, the substrate is rinsed and the surface is illuminated through a second mask, to expose additional hydroxyl groups for coupling. A second 5'-protected, 3'-*O*-phosphoramidite-activated deoxynucleoside is presented to the surface. The selective photodeprotection and coupling cycles are repeated until the desired set of products is

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: SBH, sequencing by hybridization; MeNPoc-, α -methyl-6-nitropiperonyloxycarbonyl-; DMT, 4,4'-dimethoxytrityl chloride.

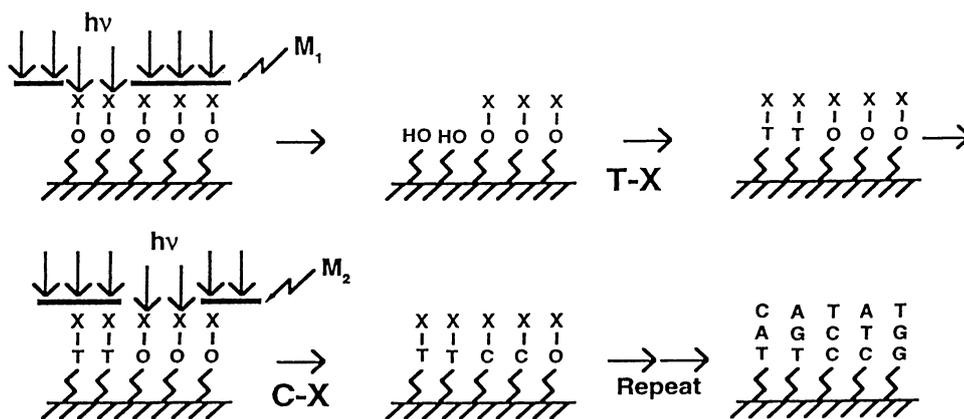


FIG. 1. Light-directed synthesis of oligonucleotides. A surface bearing photoprotected hydroxyls (X-O) is illuminated through a photolithographic mask (M_1), generating free hydroxyl groups in the photodeprotected regions. The hydroxyl groups are then coupled to a deoxynucleoside phosphoramidite (5'-photoprotected). A new mask pattern (M_2) is applied, and a second photoprotected phosphoramidite is coupled. Rounds of illumination and coupling are repeated until the desired set of products is obtained.

obtained. Since photolithography is used, the process can be miniaturized to generate high-density arrays of oligonucleotide probes. Furthermore, the sequence of the oligonucleotides at each site is known.

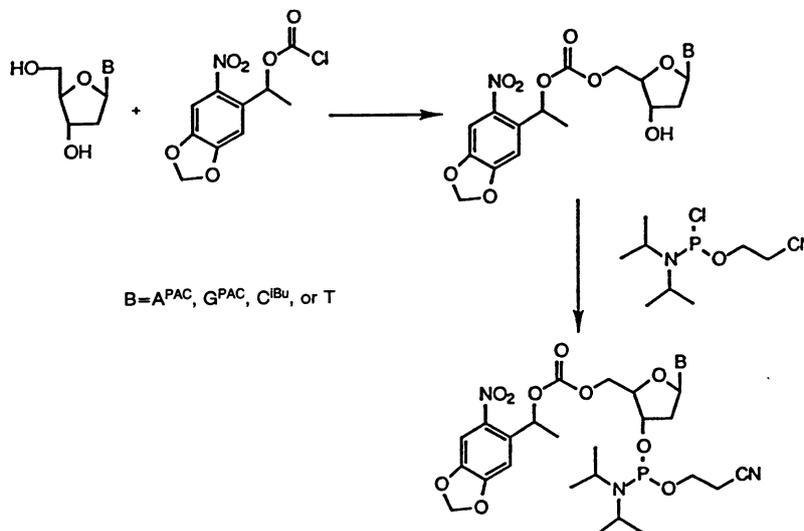
5'-Photolabile *N*-Acyl-deoxynucleoside-3'-*O*-phosphoramidites. Two factors enter into the design of photolabile 5'-hydroxyl protecting groups. Because the bases have strong $\pi-\pi^*$ transitions in the 280-nm region, the deprotection wavelength should be longer than 280 nm to avoid undesirable nucleoside photochemistry. In addition, the photodeprotection rates of the four deoxynucleosides should be similar so that light will equally effect deprotection in different illuminated synthesis sites. To meet these criteria we have developed a set of 5'-*O*-(α -methyl-6-nitropiperonyloxycarbonyl)-*N*-acyl-2'-deoxynucleosides (MeNPoc-*N*-acyl-deoxynucleosides) and measured their photokinetic behavior.

The synthetic pathway for preparing 5'-*O*-(α -methyl-6-nitropiperonyloxycarbonyl)-*N*-acyl-2'-deoxynucleoside phosphoramidites (MeNPoc-*N*-acyl-2'-deoxynucleoside phosphoramidites) is illustrated in Scheme I. In the first step, an *N*-acyl-2'-deoxynucleoside reacts with 1-(2-nitro-4,5-methylenedioxyphenyl)ethyl-1-chloroformate to yield 5'-MeNPoc-*N*-acyl-2'-deoxynucleoside. In the second step, the 3'-hydroxyl reacts with 2-cyanoethyl *N,N'*-diisopropylchlorophosphoramidite using standard procedures to yield the 5'-MeNPoc-*N*-acyl-2'-deoxynucleoside-3'-*O*-(2-cyanoethyl-*N,N'*-diisopropyl)phosphoramidites. The photoprotecting group is stable under ordinary phosphoramidite synthesis con-

ditions and can be removed with aqueous base. These reagents can be stored for long periods under argon at 4°C.

A 0.1 mM solution of each of the four deoxynucleosides—MeNPoc-dT, MeNPoc-dC^{ibu}, MeNPoc-dG^{PAC}, and MeNPoc-dA^{PAC}—was prepared in dioxane. Aliquots (200 μ l) were irradiated with 14.5 mW of 365-nm light per cm² in a narrow-path (2 mm) quartz cuvette for various times. Four or five time points were collected for each base, and the solutions were analyzed for loss of starting material at 280 nm on a Nucleosil 5-C₈ HPLC column using a mobile phase of 60% (vol/vol) in water containing 0.1% (vol/vol) trifluoroacetic acid [MeNPoc-dT required a mobile phase of 70% (vol/vol) methanol in water]. Peak areas of the residual MeNPoc-*N*-acyl-deoxynucleoside were calculated, yielding photolysis half-times of 28 s, 31 s, 27 s, and 18 s for MeNPoc-dT, MeNPoc-dC^{ibu}, MeNPoc-dG^{PAC}, and MeNPoc-dA^{PAC}, respectively. In all subsequent lithographic experiments, illumination times of 4.5 min ($9 \times t_{1/2}$ MeNPoc-dC) were used to ensure >99% removal of MeNPoc protecting groups.

The synthesis support consists of a 5.1 \times 7.6 cm glass substrate prepared by cleaning in concentrated NaOH, followed by exhaustive rinsing in water. The surfaces were then derivatized for 2 hr with a solution of 10% (vol/vol) bis(2-hydroxyethyl)aminopropyltriethoxysilane (Petrarch Chemicals, Bristol, PA) in 95% ethanol, rinsed thoroughly with ethanol and ether, dried *in vacuo* at 40°C, and heated at 100°C for 15 min. In these studies, a synthesis linker is attached by reacting derivatized substrates with 4,4'-dimethoxytrityl (DMT)-hexaethyloxy-*O*-cyanoethyl phosphoramidite.



Scheme I

In a light-directed synthesis, the overall synthesis yield depends on the photodeprotection yield, the photodeprotection contrast, and the chemical coupling efficiency. Photo-kinetic conditions are chosen to ensure that photodeprotection yields are >99%. Unwanted photolysis in normally dark regions of the substrate can adversely affect the synthesis fidelity and can be minimized by using lithographic masks with a high optical density (5 OD units) and by careful index matching of the optical surfaces.

Two different methods were developed to investigate the chemical coupling efficiencies of the photoprotected nucleosides. First, the efficiencies were measured on hexaethylene glycol derivatized control pore glass. The glycol linker was detritylated and a MeNPoc-deoxynucleoside-*O*-cyanoethyl phosphoramidite coupled to the resin. Next, a DMT-deoxynucleoside-cyanoethyl phosphoramidite (reporter amidite) was added. The reporter amidite should couple to any unreacted hydroxyl groups remaining from the first coupling reaction. Following DMT deprotection, the trityl effluents were collected and quantified by absorption spectroscopy. In this assay, the coupling efficiencies are measured assuming a high coupling efficiency of the reporter amidite. The efficiencies of the MeNPoc-deoxyribonucleoside-*O*-cyanoethyl phosphoramidites to the hexaethylene glycol linker and the 16 different dinucleotide efficiencies were measured. The values ranged between 95% and 100% in this assay and were indistinguishable from values obtained with standard DMT-deoxynucleoside phosphoramidites.

To measure the coupling efficiency of the photoprotected nucleosides directly on the glass synthesis supports, each of the four MeNPoc-amidites was coupled to a substrate. A section of the substrate was deprotected by illumination and allowed to react against a MeNPoc-phosphoramidite. A new region of the substrate was then illuminated, a fluorescent deoxynucleoside phosphoramidite (FAM-phosphoramidite; Applied Biosystems) was coupled, and the substrate was scanned for fluorescence signal. Assuming that the fluorescently labeled phosphoramidite reacts at both the newly exposed hydroxyl groups and the previously unreacted hydroxyl groups, then the ratio of fluorescence intensities between the two sites provides a measure of the coupling efficiency. This measurement of the surface chemical coupling yields efficiencies ranging between 85% and 98%.

RESULTS

Spatially Directed Synthesis of an Oligonucleotide Probe. To initiate the synthesis of an oligonucleotide probe, MeNPoc-dC^{bu}-3'-*O*-phosphoramidite was attached to a synthesis support through a linker (hexaethyloxy-*O*-cyanoethyl phosphoramidite). Regions of the support were activated for synthesis by illumination through 800 × 12800 μm apertures of a photolithographic mask. Seven additional phosphoramidite synthesis cycles were performed (with DMT-protected deoxynucleosides) to generate the sequence 3'-CGCATCCG. Following removal of the phosphate and exocyclic amine protecting groups with concentrated NH₄OH for 4 hr, the substrate was mounted in a water-jacketed thermostatically controlled hybridization chamber.

Hybridization of Targets to Surface Oligonucleotides. To explore the availability of the support-bound octanucleotide probes for hybridization, 10 nM 5'-GCGTAGGC-fluorescein was introduced into the hybridization chamber and allowed to incubate for 15 min at 15°C. The surface was then interrogated with an epifluorescence microscope (488-nm argon ion excitation). The fluorescence image of this scan is shown in Fig. 2. The fluorescence intensity coincides with the 800 × 12800 μm stripe used to direct the synthesis of the probe. Furthermore, the signal intensities are high (four times over

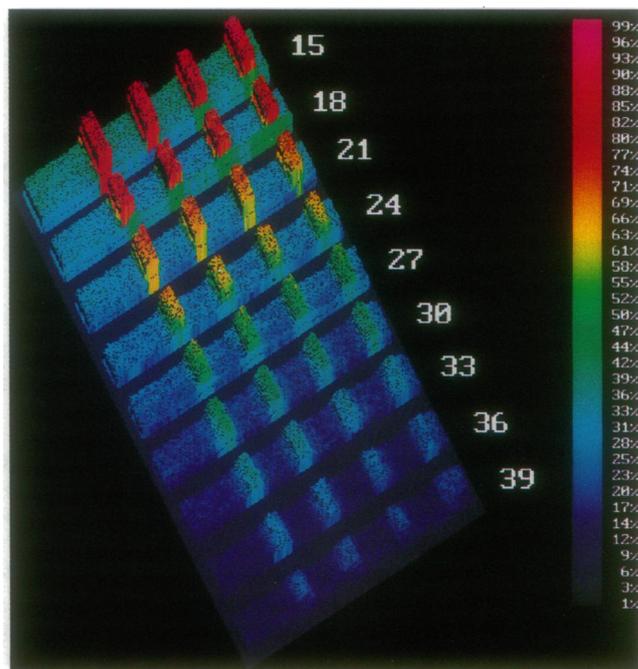


FIG. 2. Hybridization and thermal dissociation of oligonucleotides. Fluorescence scan of 5'-GCGTAGGC-fluorescein hybridized to complementary probes. The substrate surface was scanned with a Zeiss Axioscop 20 microscope using 488-nm argon ion laser excitation. The fluorescence emission above 520 nm was detected using a cooled photomultiplier (Hamamatsu 934-02) operated in photon-counting mode. The signal intensity is indicated on the color scale shown to the right of the image. The temperature is indicated to the right of each panel in °C.

the background of the glass substrate), demonstrating specific binding of the target to the probe.

The melting behavior of the target-probe complex was investigated by increasing the temperature in the hybridization chamber. After 10 min of incubation at each temperature, no significant changes in the fluorescence were observed, and the fluorescence intensities were recorded. The duplex melted in the temperature range expected for the sequence under study. The probes were stable to temperature denaturation of the target-probe complex as demonstrated by rehybridization of target DNA.

Sequence Specificity of Target Hybridization. To demonstrate the sequence specificity of probe hybridization, two different sequences were synthesized in 800 × 12800 μm stripes. Fig. 3A identifies the location of the two probes. The probe 3'-CGCATCCG was synthesized in stripes 1, 3, and 5. The probe 3'-CGCTTCCG was synthesized in stripes 2, 4, and 6. Fig. 3B shows the results of hybridizing 5'-GCGTAGGC-fluorescein to the substrate at 15°C. Although the probes differ by only one internal base, the target hybridizes specifically to its complementary sequence (~500 counts above background in stripes 1, 3, and 5) with little or no detectable signal in positions 2, 4, and 6 (~10 counts). Fig. 3C shows the results of hybridization with targets to both sequences. The signal in all positions in Fig. 3C illustrates that the absence of signal in Fig. 3B is due solely to the instability of the single base hybridization mismatch. Although the targets are present in equimolar concentrations, the ratios of signals in stripes 2, 4, and 6 in Fig. 3B are approximately 1.6 times higher than the ratios of signals in regions 1, 3, and 5, presumably due to slight differences in melting temperature values for the two duplexes. The duplexes were dissociated by raising the temperature to 45°C for 15 min, and the hybridizations were repeated in the reverse

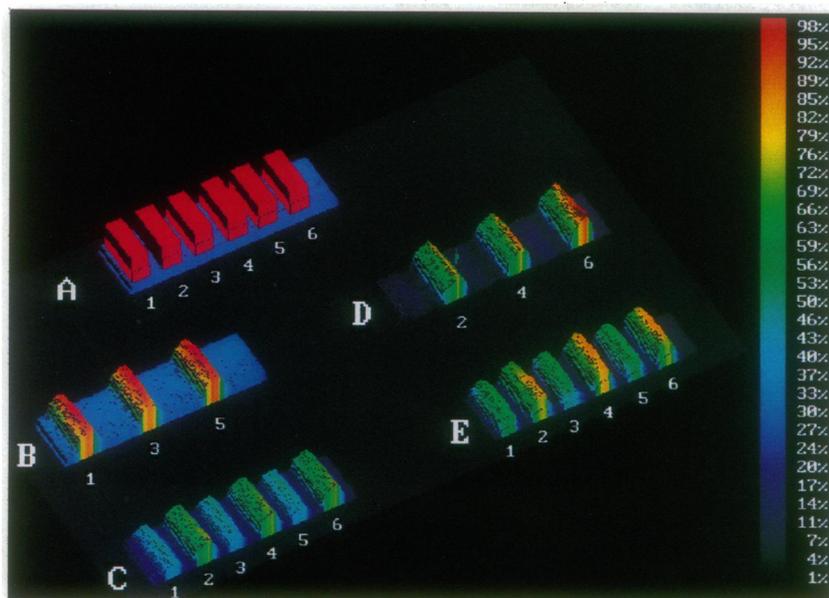


FIG. 3. Sequence specificity of hybridization. (A) Index of the probe composition at each synthesis site. 3'-CGCATCCG was synthesized in stripes 1, 3, and 5, and 3'-CGCTTCCG was synthesized in stripes 2, 4, 6. (B) Fluorescence image showing hybridization of substrate with 10 nM 5'-GCGTAGGC-fluorescein. Hybridization was performed in 6× SSPE (1 M NaCl/66 mM sodium phosphate/6 mM EDTA, pH 7.4)/0.1% Triton X-100 at 15°C for 15 min. (C) Hybridization with 10 nM 3'-GCGAAGGC added to the hybridization solution of B. (D) After high-temperature dissociation of fluoresceinated targets from C, the substrate was incubated with 10 nM 3'-GCGAAGGC at 15°C for 15 min. (E) Hybridization with 10 nM 3'-GCGTAGGC added to the hybridization solution of D.

order (Fig. 3 D and E), demonstrating specificity of hybridization in the reverse direction.

Combinatorial Synthesis of a Probe Matrix. In a light-directed synthesis, the location and composition of products depend on the pattern of illumination and the order of chemical coupling reagents (6). Consider the synthesis of 256 tetranucleotides, as illustrated in Fig. 4. Mask 1 activates one-fourth of the substrate surface for coupling with the first of 4 nucleosides in the first round of synthesis. In cycle 2, mask 2 activates a different quarter of the substrate for coupling with the second nucleoside. The process is continued to build four regions of mononucleotides. The masks of round 2 are perpendicular to those of round 1, and each cycle generates four new dinucleotides. The process continues through round 2 to form 16 dinucleotides as illustrated in Fig. 4. The masks of round 3 further subdivide the synthesis regions so that each coupling cycle generates 16 trimers. The subdivision of the substrate is continued through round 4 to form the tetranucleotides. The synthesis of this probe matrix can be compactly represented in polynomial notation (6) as $(A + C + G + T)^4$. Expansion of this reaction polynomial identifies the 256 tetranucleotides.

The potential of light-directed combinatorial synthesis can be appreciated by the combinatorial synthesis of the probe matrix shown in Fig. 5A. The polynomial for this synthesis is given by 3'-CG(A + G + C + T)⁴CG. The synthesis map is given in Fig. 5B. All possible tetranucleotides are synthesized flanked by CG at the 3' and 5' ends. Hybridization of the target 5'-GCGGCGGC-fluorescein to this array at 15°C yields

the 3'-CGCCGCCG complementary probe as the most intense position (2698 counts). Significant intensity is also observed for the following mismatches: 3'-CGCAGCCG (554 counts), 3'-CGCCGACG (317 counts), 3'-CGCCGTCG (272 counts), 3'-CGACGCCG (242 counts), 3'-CGTCGCCG (203 counts), 3'-CGCCCCCG (180 counts), 3'-CGCTGCCG (163 counts), 3'-CGCCACCG (125 counts), and 3'-CGCCTCCG (78 counts).

DISCUSSION

Arrays of oligonucleotides can be efficiently generated by light-directed synthesis and can be used to determine the identity of DNA target sequences. As shown here, an array of all tetranucleotides was produced in 16 cycles requiring 4 hr to complete. Because combinatorial strategies are used (6), the number of compounds increases exponentially while the number of chemical coupling cycles increases only linearly. For example, expanding the synthesis to the complete set of 4⁸ (65,536) octanucleotides will add only 4 hr to the synthesis for the 16 additional cycles. Furthermore, combinatorial synthesis strategies can be implemented to generate arrays of any desired composition (6). For example, since the entire set of dodecamers (4¹²) can be produced in 48 photolysis and coupling cycles (bⁿ compounds requires b × n cycles), any subset of the dodecamers (including any subset of shorter oligonucleotides) can be constructed with the correct lithographic mask design (6) in 48 or fewer chemical coupling steps. The number of compounds in an array is limited only

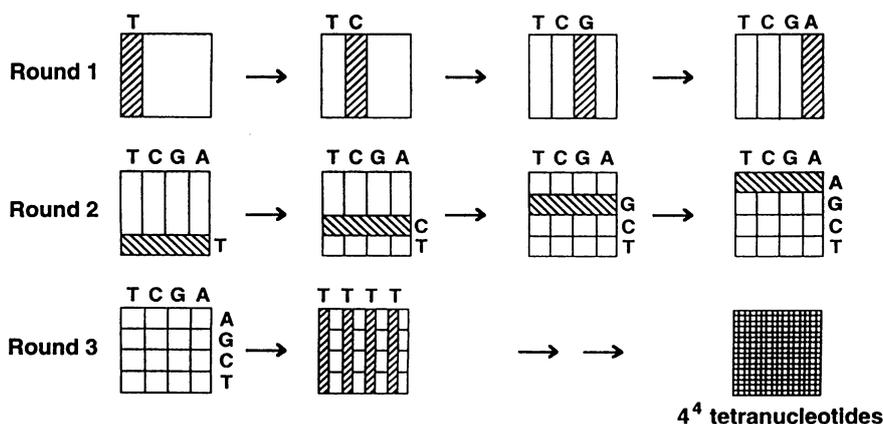


FIG. 4. Combinatorial synthesis of 4⁴ tetranucleotides. In round 1, one-fourth of the synthesis area is activated by illumination through mask 1 for coupling of the first MeNPoc-nucleoside (A in this case). In cycle 2 of round 1, mask 2 activates a different one-quarter section of the synthesis substrate and a different nucleoside (T) is coupled. Further lithographic subdivisions of the array and chemical couplings generate the complete set of 256 tetranucleotides as described in the text.

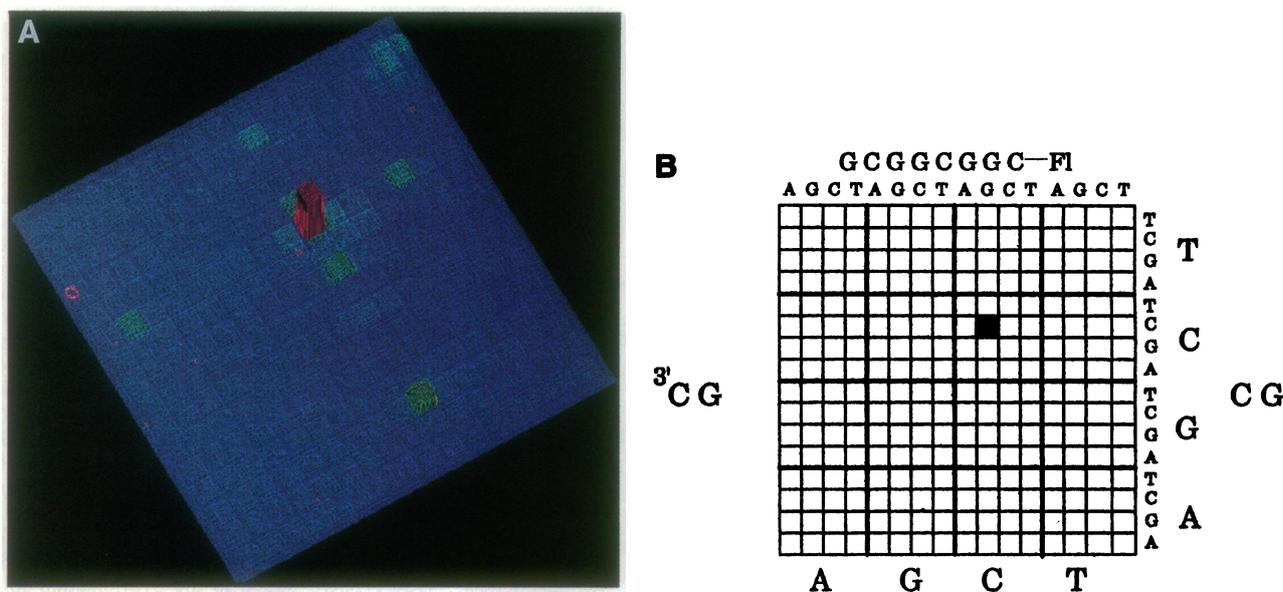


FIG. 5. Hybridization to a matrix of 256 octanucleotides. (A) Fluorescence image of a 256-octanucleotide array following hybridization with 10 nM 5'-GCGGCGGC-fluorescein in $6\times$ SSPE/0.1% Triton X-100 for 15 min at 15°C. (B) Matrix decoder for the array. The matrix synthesis is represented by the polynomial $3'-CG(A + G + C + T)^4CG$. The site containing the complementary sequence 3'-CGCGCCCG is highlighted for reference.

by the density of synthesis sites and the overall array size. Recent experiments have demonstrated hybridization to probes synthesized in 25- μ m sites. At this resolution, the entire set of 65,536 octanucleotides can be placed in an array measuring 0.64-cm square, and the set of 1,048,576 dodecanucleotides requires only a 2.56-cm array.

Genome sequencing projects will ultimately be limited by DNA sequencing technologies. Current sequencing methodologies are highly reliant on complex procedures and require substantial manual effort (8, 9). SBH has the potential for transforming many of the manual efforts into more efficient and automated formats. Light-directed synthesis is an efficient means for enabling SBH and for the large-scale production of miniaturized arrays.

Oligonucleotide arrays are not limited to primary sequencing applications. Because single base changes cause multiple changes in the hybridization pattern, the oligonucleotide arrays will provide a powerful means to check the accuracy of previously elucidated DNA sequence or to scan for changes within a sequence. In the case of octanucleotides, a single base change in the target DNA results in the loss of eight complements and generates eight new complements. Matching of hybridization patterns may be useful in resolving sequencing ambiguities from standard gel techniques or for rapidly detecting DNA mutational events.

The potentially very high information content of light-directed oligonucleotide arrays will change genetic diagnostic testing. Sequence comparisons of hundreds to thousands of different genes will be assayed simultaneously instead of the current one, or few at a time format. Custom arrays can also be constructed to contain genetic markers for the rapid identification of a wide variety of pathogenic organisms.

Oligonucleotide arrays can also be applied to study the sequence specificity of RNA- or protein-DNA interactions. Experiments can be designed to elucidate specificity rules of non-Watson-Crick oligonucleotide structures or to investigate the use of novel synthetic nucleoside analogs for antisense or triple helix applications. Suitably protected RNA monomers may be employed for RNA synthesis. The oligonucleotide arrays should find broad application deducing the thermodynamic and kinetic rules governing formation and stability of oligonucleotide complexes.

We are grateful to L. Stryer, P. Schultz, and F. Pease for their critical reading and helpful comments on the manuscript. A.C.P. is a National Institutes of Health postdoctoral fellow (HG 00060). This work was sponsored in part by National Institutes of Health (HG 00813) and Department of Energy (DE FG03 92ER) grants to S.P.A.F.

1. Lysov, Yu. P., Florentiev, V. L., Khorlyn, A. A., Khrapko, K. R., Shick, V. V. & Mirzabekov, A. D. (1988) *Dokl. Akad. Nauk SSSR* **303**, 1508-1511.
2. Bains, W. & Smith, G. C. (1988) *J. Theor. Biol.* **135**, 303-307.
3. Drmanac, R., Labat, I., Brukner, I. & Crkvenjakov, R. (1989) *Genomics* **4**, 114-128.
4. Strezoska, Z., Paunesku, T., Radosavljevic, D., Labat, I., Drmanac, R. & Crkvenjakov, R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10089-10093.
5. Drmanac, R., Drmanac, S., Strezoska, Z., Paunesku, T., Labat, I., Zeremski, M., Snoddy, J., Funckhouser, W. K., Koop, P., Hood, L. & Crkvenjakov, R. (1993) *Science* **260**, 1649-1652.
6. Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Tsai Lu, A. & Solas, D. (1991) *Science* **251**, 767-773.
7. Southern, E. M., Maskos, U. & Elder, J. K. (1992) *Genomics* **13**, 1008-1017.
8. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
9. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560-564.