Fractionation, phosphorylation and ligation on oligonucleotide microchips to enhance sequencing by hybridization

Svetlana Dubiley¹,+ , Eugene Kirillov¹,+ , Yuri Lysov¹,+ and Andrei Mirzabekov¹,2,*

Joint Human Genome Program: ¹Engelhardt Institute of Molecular Biology, 32 Vavilov Str., Moscow, 117984, Russia and ²Center for Mechanistic Biology and Biotechnology, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA

Received April 14, 1997; Accepted May 1, 1997

ABSTRACT

Oligonucleotide microchips are manufactured by immobilizing presynthesized oligonucleotides within 0.1 × 0.1 × 0.02 mm or 1 × 1 × 0.02 mm polyacrylamide gel pads arranged on the surface of a microscope slide. The gel pads are separated from each other by hydrophobic glass spacers and serve as a kind of 'microtest tube' of 200 nl or 20 nl volume, respectively. Fractionation of single-stranded DNAs is carried out by their hybridization with chip pads containing immobilized 10mers. DNA extracted separately from each pad is transferred onto a sequencing chip and analyzed thereon. The chip, containing a set of 10mers, was enzymatically phosphorylated, then hybridized with DNA and ligated in a site-directed manner with a contiguously stacked 5mer. Several cycles of successive hybridization–ligation of the chip-bound 10mers with different contiguously stacked 5mers and hybridized with DNA were carried out to sequence DNA containing tetranucleotide repeats. Combined use of these techniques show significant promise for sequence comparison of homologous regions in different genomes and for sequence analysis of comparatively long DNA fragments or DNA containing internal repeats.

INTRODUCTION

The programs to map and sequence the human genome and genomes of other organisms have been rapidly expanding (1). The genome studies are based on a number of operations which are repeated many times for many samples, such as enzymatic reactions, PCR amplification, cloning, gel electrophoresis, hybridization, sequencing, etc. Many of these procedures are costly and labor intensive, which has stimulated the search for automated procedures, robots, high-density arrays and a parallel processing of many DNA samples.

Sequencing is still a major bottleneck in the genome programs, requiring a significant increase in the efficiency of conventional gel sequencing methods and the development of basically new sequencing procedures (1). One of these procedures, sequencing by hybridization, is based on the hybridization of DNA to a large array of immobilized oligonucleotides, for example, to all 65 536 octamers. Identification of all 5mers present in a DNA molecule may enable one to identify overlapping oligonucleotides and arrange them in their specific order, thus reconstituting the DNA sequence (2–6). Oligonucleotide arrays have been manufactured by direct synthesis of oligonucleotides on a solid support (6–8) or by immobilization of presynthesized oligomers (4,9–12).

In our method of sequencing by hybridization, the oligonucleotides are immobilized within the micropads of polyacrylamide gel (2,4,9). Hybridization of these oligonucleotides with DNA in the presence of 5mers contiguously stacked with the immobilized oligonucleotides (contiguous stacking hybridization) has been proposed to increase the efficiency of sequencing (4,9,13,14). These microchips have been applied for sequence analysis (12), mutation analysis (15) and identification of microorganisms (16).

Here we describe the use of oligonucleotide microchips for DNA fractionation, site-directed oligonucleotide phosphorylation and ligation, as well as for hybridization–ligation 'walking'. These new applications of the chip can facilitate the sequence analysis of long DNA, for example, cosmid DNA containing oligonucleotide repeats. The use of these procedures may significantly simplify nucleotide sequence comparison of similar genes and genomes.

MATERIALS AND METHODS

Oligonucleotide synthesis. PCR amplification of a 421 bp fragment of human β-globin gene, and the manufacturing of oligonucleotide chips were described earlier (12).
Fractionation of DNA on an array of gel-immobilized oligonucleotides

Three nested fragments F1, F2 and F3 (Fig. 1A), were amplified separately with fluorescently labeled sense primers (5′-TMR-TGGAAGCCACACCTAGG-3′ for F1, 5′-TMR-A-GAAATGCTGCGTTACTGCC-3′ for F2, 5′-TMR-TGGAAGCACTCTTTTG-3′ for F3) and antisense primers (5′-ACCTTTTATGCCCCAGCCCC-3′ for F1, 5′-CTTTGATAC- CACCTTGCCC-3′ for F2, 5′-AAAATAGACCAATAGGCA- GAGAG-3′ for F3). Reaction mixtures contained 50 ng 421 bp DNA target, 1 U AmpliTaq polymerase and the corresponding primers. Amplifications were carried out in 30 cycles: 94°C for 30 s, 50°C for 40 s and 72°C for 20 s. Then 3 μl of each PCR mixture was used for single primer reamplification with the corresponding fluorescently labeled sense primers, and 40 cycles of amplification were carried out under the conditions described above. Fifty microliters of each PCR mixture was precipitated separately with 500 μl of 0.2 M LiClO₄ in acetone, and the remaining 50 μl of each mixture was mixed in a single tube and also precipitated. The pellets were washed with acetone and dissolved in hybridization buffer (1 M NaCl, 1 mM EDTA, 10 mM sodium phosphate, pH 6.8).

Two microliters of fluorescently labeled F1, F2 and F3, and their mixture were hybridized separately with the fractionation chip, which contained three 10mers (f1, f2 and f3; 15 pmol of each) immobilized separately within the 1×1×0.02 mm gel pads (Fig. 1B). To prevent evaporation, the chip was placed on wet paper in a Petri dish. After incubation at 4°C for 12 h, the chip was washed with hybridization buffer at 4°C for 5 min, followed by washing at 15°C for 5 min to remove non-hybridized DNA.

DNA fragments were eluted separately from each of three F1+2+3 pads of the fractionation chip at 40°C with 3 μl hybridization buffer and hybridized under the conditions described above with a microchip containing nine 10mers (Fig. 1C) for DNA sequence analysis (sequencing microchip). Elution of fractionated DNA can be also carried by blotting. A fractionation chip was heated (by a Peltier element) to 40°C and was placed onto a sequencing chip cooled (by another Peltier element) to 4°C. The DNA of the duplexes melted at the high temperature of the fractionation chip diffused into the sequencing chip, whereon it was captured by hybridization at the low temperature (not shown). Fluorescence hybridization images of the fractionation and the sequencing microchips were monitored with a fluorescence microscope equipped with a CCD camera and appropriate software (12).
Phosphorylation and ligation on oligonucleotide chips

Fifty picomoles of synthetic oligodeoxyribonucleotide (5′-ATA-CCACCT-3′) were phosphorylated in 10 µl reaction mixture containing 1× PN kinase buffer (used as suggested by Epicentre Technologies, USA), 15 µCi [γ-32P]ATP, 500 pmol ATP and 0.5 U T4 polynucleotide kinase (Epicentre Technologies, USA) at 37°C for 60 min. The 32P-labeled 10mer was purified from unincorporated label with a Bio-Spin 6 Chromatography Column (Bio-Rad Laboratories, USA) and immobilized within gel pad I (1 × 1 × 0.02 mm). The non-phosphorylated oligomer of the same sequence (Fig. 2A) was immobilized within the II, III and IV pads of the chip (5 pmol/pad). Immobilization was carried out at 20°C for 12 h at 100% humidity. The chip was washed with washing buffer (0.2 M NaCl, 0.2 mM EDTA, 2 mM sodium phosphate, pH 6.8) at 37°C for 1 h and then rinsed with water. The chip was dried and autoradiographed with Kodak Scientific Imaging Film X-OMAT (Eastman Kodak Co., USA) (Fig. 2B1).

One microliter of phosphorylation mixture (1.5 µCi [γ-32P]ATP, 50 pmol ATP and 0.05 U T4 polynucleotide kinase in 1× PN kinase buffer) was added to gel element II. As a control, either the same mixture (excluding kinase) was added to pad III, or 1 µl 1× PN kinase buffer was added to pad IV. Enzymatic phosphorylation was carried out at 37°C for 3 h at 100% humidity. The chip was washed first with the washing buffer at 37°C for 1 h, and then with water, followed by drying and autoradiography. A 2.5 µl volume of ligation mixture (containing 10 pmol single-stranded (ss) DNA, 5′-TGAGTGTACGTCCAGTCA TCACA TACA TACA TACA TACAA TT-3′, complementary to the immobilized 10mer, 50 pmol of fluorescently labeled 5mer, 5′-HEX-ccttg-3′, stacked to the immobilized 10mer, 1 mM ATP and 0.1 U T4 DNA ligase [Epicentre Technologies] in 1× T4 DNA ligase buffer [Epicentre Technologies]) was added to pads II and III. As a control, 2.5 µl of the same mixture, excluding ligase, was added to the pads I and IV. Ligation was carried out at 4°C for 5 h; then the chip was washed with the washing buffer at 10°C for 5 min.

Stepwise hybridization–ligation on the chip

Five pentamers P1 (5′-TGATG-3′), P2 (5′-TGTAT-3′), P3 (5′-ATGTG-3′), P4 (5′-TATGT-3′) and P5 (5′-AATTG-3′), a 10mer (5′-TGATGACTGG-3′), and ssDNA (5′-TGAGTGTACGTCCAGTCA TCACA TACA TACA TACA TACAA TT-3′) were synthesized. Pentamers of the same sequence labeled at the 5′-end with fluorescein were also synthesized.

One picomole of 10mer and 40 pmol each of the P1–P4 pentamers were phosphorylated separately with a 10-fold excess of ATP and 1 U T4 polynucleotide kinase, as described above. The phosphorylated 10mer was immobilized within four gel pads (100 fmol per 0.1 × 0.1 × 0.02 mm pad) of the chip. Five cycles of successive contiguous stacking hybridization followed by ligation were carried out under the same conditions. The first hybridization was carried out in 10 µl hybridization buffer containing 1 µM fluorescently labeled pentamer P1 and 1 µM complementary ssDNA. A 10 µl drop of the hybridization solution was placed on the chip, covered by a cover-glass (Corning, USA) over 0.1-mm-thick spacers, and incubated for...
Fractionation chip. Probability calculations

Let us presume that all four bases can occur with the same 1/4 probability in any position of an analyzed DNA molecule. Then, the sequence of a DNA segment of length $m$ can be identical to the sequence of a fractionation chip oligonucleotide of the same length $m$ with the probability $1/4^m$. Accordingly, the probability that these sequences are different is $(1 - 1/4^m)$. In DNA of length $l$, the total number of $m$-mers is $(l-m+1)$, and the probability $q^m$ that all the $m$-mers differ from a selected oligonucleotide is $q^m = (1 - 1/4^m)^l(m-1)$. Accordingly, the probability $q^m$ that at least one DNA segment has the same sequence as the oligonucleotide is $q^m = 1 - (1 - 1/4^m)^l(m-1)$.

Some DNA sequences in the mixture can be hybridized with several oligonucleotides of the chip. On the other hand, some immobilized oligonucleotides will be hybridized simultaneously with several DNAs of the mixture.

Let us assume that DNA upon digestion by a restriction endonuclease produces $K$ fragments of similar length. With a probability $P_{21} = 1 - (q^m)^K$, each immobilized oligonucleotide will hybridize with at least one fragment. A chip oligonucleotide is hybridized only with one chosen fragment and thus separates it from the others with a probability $P_{2} = q^m \cdot (q^m)^{K-1}$. The probability $P_{2}$ that an oligonucleotide is not hybridized with the chosen fragment or it hybridizes with two or more fragments (there is no isolation of individual fragments) is $P_{2} = 1 - q^m \cdot (q^m)^{K-1}$. Immobilized oligonucleotides can be selected of non-correlated nucleotide sequences. Then, the probability $P$ that hybridization of the fragment mixture with $N$ oligonucleotides does not lead to isolation of the individual fragment would be $P_{2} = (1 - q^m \cdot (q^m)^{K-1})^N$. The minimal number of oligonucleotides in the chip needed for isolating every fragment from the mixture with a certain probability can be calculated from this equation.

The number $n$ of the oligonucleotides within a chip with $N$ immobilized oligonucleotides that are hybridized with any DNA fragment is $n = N \cdot P_{21}$. The ratio of oligonucleotides hybridized with one fragment ($n_1$) to oligonucleotides hybridized with any number of fragments is $n_1/n = P_{1}/P_{21}$.

Due to the uneven distribution of restriction sites along DNA, deviation in the DNA sequence from random, and non-ideal discrimination of perfect duplexes from the duplexes containing some mismatches, the real number of oligonucleotides in the fractionation chip should be somewhat larger.

RESULTS

The array of gel-immobilized oligonucleotides

The oligonucleotide array is formed of polyacrylamide gel pads (gel matrix) fixed on a glass slide (12). The pads are separated from one another by hydrophobic glass spacers. The spacers prevent cross-talk between the tiny droplets of water solutions placed on adjacent pads. Therefore, the gel matrix can be used as an array of microtest tubes to carry out chemical, enzymatic or physicochemical procedures separately in each gel pad. The gel matrices of two sizes were prepared: 1 × 1 × 0.02 mm, 20 nl in volume, for manual loading with the micropipette, as well as 0.1 × 0.1 × 0.02 mm, 200 pl in volume, for robotics loading (12) of oligonucleotides onto a specific gel pad. Equipment for manual loading on up to 100.1 × 0.1 × 0.02 mm gel pads has been developed (19). To prevent the evaporation of the tiny volumes of water solutions from the gel pads during chemical or enzymatic procedures, microchips were either placed in 100% humidity, covered with a glass cover slip or an oil layer (12). Chemically active groups were introduced into both the polyacrylamide gel and the oligonucleotides to cause their covalent crosslinking (9,17). Oligonucleotide microchips were fabricated with a robot that applies ~1 nl of activated oligonucleotide solution onto a specified 0.1 × 0.1 × 0.02 mm gel pad once or several times with a robot pin. Such oligonucleotide chips can sustain 15–50 rounds of hybridization at 0–50°C.

Fractionation and sequence analysis on oligonucleotide arrays

Complementary interactions are effectively used for affinity fractionation of nucleic acids by hybridization with complementary oligonucleotides immobilized in a solid support. Isolation of polyA-containing mRNA on immobilized poly(dT) is an example of this well-developed procedure.

An array of gel-immobilized 10mers was tested for fractionation of single-stranded DNA fragments of the human β-globin gene (Fig. 1A). The array consisted of 12 gel elements (1 × 1 × 0.02 mm) containing three immobilized fractionating 10mers (f1, f2 and f3) arranged in three rows (Fig. 1B). Each of three fluorescently labeled DNA fragments (F1, F2 and F3) were complementary to the f1, f2 and f3 10mers, respectively. The three fragments and a mixture thereof (F1+2+3) were applied on the gel pads of the fractionation chip. Hybridization was carried out at 4°C followed by washings, first at 4°C and then at 15°C. Figure 1B (see columns F1, F2 and F3) shows that each DNA fragment was hybridized with only a complementary 10mer, and their mixture with all three 10mers.

The hybridized DNA was eluted separately from each of the three F1+2+3 pads. Each fraction of extracted DNA was hybridized with the chip prepared for sequence analysis (sequencing chip) (Fig. 1C). The chip contained nine 10mers (six ‘s’; and three ‘f’ 10mers; Fig. 1A) that were complementary to the DNA fragments. Figure 1C shows that each DNA isolated from the F1+2+3 mixture on each fractionating 10mer hybridizes only with the corresponding set of 10mers, and thus it can be unambiguously identified. The cross-hybridization of the 10mers with DNAs of the other two fractions was negligible. These results demonstrate the specificity of DNA separation on the fractionation chip and its compatibility with sequence analysis on the sequencing chip.
DNA fractionation on the oligonucleotide array containing specified or randomly selected oligonucleotides can be used for isolating specific fragments from DNA digested with restriction endonucleases. To separate the fragments from each other, the fractionation chip should contain oligonucleotides that are unique for every fragment. Probability calculations allow a rough estimate of the fractionation efficiency and the number of oligonucleotides of various lengths immobilized in the chip to isolate specific DNA fragments from mixtures of varying complexity.

Table 1 shows the average number $K$ of fragments ($\sim$250 and 4000 bases long) produced by digestion of a random DNA sequence having $\sim$50% A/T content from a plasmid ($\sim$4 kb), a cosmid ($\sim$30 kb) and BAC ($\sim$100 kb) with restriction endonucleases specific to four and six bases. Table 1 lists also the calculated number of oligonucleotides, $N$, of length $m$ in an array that would be needed to fractionate the mixture of $K$ random sequence DNA fragments of length $l$ with 90, 95 or 99% probability. $N$ signifies the minimal number of oligonucleotides in the fractionation chip wherein each fragment is hybridized, within the specified probability, to one and only one oligonucleotide. The numbers $n_1$ and $n$ designate the number of oligonucleotides among $N$ that bind only one or one or more fragments, respectively. The ratio $n_1/n$ represents the efficiency of the chip, indicating the share of oligonucleotides that bind only one fragment from the mixture.

For example, to fractionate 120 250-nucleotide-long fragments derived from cosmid DNA (30 000 bp in length) with 99% probability, one will have to use chips bearing $\sim$1911 non-correlated 8mers. Among them, 697 8mers will bind DNA, and only 79% of them (i.e., 550 octamers) will isolate a single fragment from the mixture and thus participate in the fractionation.

### Site-directed phosphorylation and ligation of oligomers on the chip

The experiments on phosphorylation and stacked oligonucleotide ligation within defined pads of the chip are outlined in Figure 2A. A gel-immobilized 10mer was enzymatically phosphorylated with polynucleotide kinase, hybridized with DNA, hybridized with a stacked 5mer and finally ligated. The 10mer lacking the 5'-phosphate group cannot be ligated. Phosphorylation thus provides an additional step to prevent unwanted ligation.

The results of site-directed phosphorylation and ligation of oligonucleotides on specified pads of the chip are shown in Figure 2B. The results of site-directed phosphorylation and ligation of oligomers on specified pads of the chip are shown in Figure 2B. T4 polynucleotide kinase (140 kDa) and T4 DNA ligase (54 kDa) diffused into the 4% polyacrylamide gel of the chip pads and catalyzed the phosphorylation and ligation of immobilized oligonucleotides. A 32P-labeled 10mer was immobilized within element I and a non-phosphorylated 10mer of the same nucleotide sequence was attached to the II, III and IV pads of the array (Fig. 2B1). The mixture of the kinase and [$\gamma$-32P]ATP in a buffer was added by a pipette onto the II element only. The mixture (not containing kinase) was added to the III element as a control. After incubation of the array at 37°C for 3 h, the 32P label was detected only on pad II (Fig. 2B2). No crosstalk was observed between elements II and III or II and IV, which shows that phosphorylation of the gel-immobilized oligonucleotide was efficient and site-directed.

At this point, a complementary 21-base DNA, a fluorescently labeled 5mer and DNA ligase were added to the phosphorylated and non-phosphorylated 10mers in the II and III elements, respectively. The same solution (not containing ligase) was added to the control elements I and IV containing phosphorylated and non-phosphorylated 10mers. After completion of ligation at 4°C for 5 h and subsequent washing, the fluorescence label was observed only in element II (Fig. 2B3).

### Successive hybridization–ligation on the chip

Figure 3A shows the results of hybridization–ligation of synthetic ssDNA containing five tetranucleotide 3'-ACAT-5' repeats. The successive hybridization–ligation ‘walking’ of five 5mers along this DNA hybridized to a chip-bonded 10mer was carried out to measure the number of repeats. The 10mer complementary to the DNA was immobilized in four adjacent gel pads of the chip (Fig. 3C). Five fluorescently labeled and unlabeled 5mers (P1, P2, P3, P4 and P5) are capable of stacking with each other, and P1 was stacked with the 10mer upon hybridization with complementary DNA (Fig. 3A).

The chip was first hybridized with unlabeled DNA and the fluorescently labeled 5mer P1 (Fig. 3B, first column). P1 and the 10mer form a stable contiguous 15 bp duplex. The 5mer is

Table 1. The complexity of an oligonucleotide chip needed for fractionating various numbers of DNA fragments of different length

| DNA length | DNA fragments | Oligomer length | The number of oligonucleotides in array
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>$l$</td>
<td>$K$</td>
<td>$m$</td>
</tr>
<tr>
<td>4000</td>
<td>256</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>191</td>
<td>41</td>
<td>248</td>
</tr>
<tr>
<td>30 000</td>
<td>256</td>
<td>120</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>956</td>
<td>348</td>
<td>1243</td>
</tr>
<tr>
<td>30 000</td>
<td>4096</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>58</td>
<td>22</td>
<td>76</td>
</tr>
<tr>
<td>100 000</td>
<td>4096</td>
<td>25</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>216</td>
<td>69</td>
<td>281</td>
</tr>
</tbody>
</table>

An $L$ bp long random sequence DNA of a plasmid ($\sim$4 kb), cosmid ($\sim$30 kb) or BAC ($\sim$100 kb) is fragmented by a 4 or 6 bp specific restriction endonuclease to produce, on average, $K$ number of fragments $\sim$0.25 or 4 kb in length ($l$), respectively. The oligonucleotide chip should contain $N$ number of immobilized oligonucleotides of length $m$ to be capable of fractionating most of the ssDNA fragments. Each fragment of the mixture could be isolated with a probability $P$ on an oligonucleotide of the array without admixture of others. In such an array only $n_1$ or $n$ among $N$ oligonucleotides bind at least one, or one or more DNA fragments, respectively.
Figure 3. Sequencing of tetranucleotide repeats by successive hybridization–ligation ‘walking’. The chip contains a 10mer immobilized in parallel within four 0.1 × 0.1 × 0.02 mm gel pads. The 10mer was hybridized successively with DNA containing ACAT repeats and with fluorescently labeled pentamers (P1, P3, P4 and P5). Then the labeled 5mers were washed off, and unlabeled P1 was hybridized and ligated. After the ligation, the chip 15mer was hybridized successively with labeled P1 and P2. After washing off the labeled 5mers, unlabeled P2 was ligated with the 15mer. The same steps were repeated with P3 and P4. Finally, the ligated chip 30mer was hybridized with labeled P5. (A) The sequence of DNA, the chip-immobilized 10mer and the five 5mers (P1–P5) used in the hybridizations and ligations. (B) The relative fluorescence intensities (measured in arbitrary units, AU) of the labeled 5mers (P1, P3, P4 and P5) hybridized with the duplex of the DNA and the immobilized 10mer following four successive ligation steps of the 10mer with P1, P2, P3 and P4. (C) The pattern of the 5mer hybridization as described in B: (1) P1 is hybridized with the unligated chip 10mer; (2) P5 is hybridized with the unligated 10mer; (3) P5 is hybridized with the 30mer chip after successive ligation with P1, P2, P3 and P4.

stabilized in the duplex by the stacking interactions of adjacent bases at the terminal positions of P1 and the 10mer. The other 5mers (P2, P3, P4 and P5) are not prone to hybridization due to the lack of stabilizing stacking interactions. The labeled 5mers were washed off the chip and unlabeled 5mer P1 was added and ligated to the 10mer in the duplex with DNA. This step extends the 10mer to a complementary 15 bp duplex. In the second cycle of ‘walking’, labeled 5mer P2 was hybridized with the duplex of DNA and the chip 15mer (Fig. 3B, column P1). However, P1 is also hybridized to this duplex due to the incomplete ligation at the previous cycle. Upon hybridization and washing, the unlabeled P2 was ligated to the 15mer to form a 20mer chip. Two additional cycles of hybridization and ligation ‘walking’ with P3 and P4 correspondingly elongate the 20mer chip to 25- and 30mers. The final hybridization of the 30mers formed after four ligations is carried out with DNA and labeled P5 that does not contain the repeat sequence. This allows us to assess the 5′-flanking sequence and the length of the tetranucleotide repeat as 4 × 5 = 20 bases.

DISCUSSION

Porous gels have a 100–1000 times higher capacity for threedimensional immobilization of oligonucleotides as compared with two-dimensional immobilization, for example, on a glass surface (12). The immobilization capacity of a 1 × 1 × 0.02 mm gel pad is ~1 nmol (17), which may allow isolation of DNA in quantities sufficient for performing sequencing and other analytical procedures.

It is of interest to consider the successive use of fractionation and sequencing chips for the analysis of comparatively long DNA fragments. For example, 30 000-bp-long cosmid DNA can be digested with a 4 bp specific restriction endonuclease into ~120 fragments of 250 bp (Table 1). Cloning or other procedures can be used to generate single-stranded DNA or RNA from these fragments. For example, a T7 promoter containing DNA can be ligated to one end of each DNA fragment in the digestion mixture. The T7 RNA polymerase reaction will generate ~120 different single-stranded RNAs ~250 nucleotides in length. The 250-base RNA fragments obtained in this way (15,18) can be hybridized to the chip oligonucleotides (19). In order to isolate every RNA fragment from the mixture with a 99% probability, one needs to use a fractionation chip containing ~2000 heptamers or octamers (Table 1). After separation from each other, these RNA fragments can be sequenced individually by conventional methods or analyzed on sequencing chips. This procedure avoids intermediate subcloning of cosmid DNA fragments and can be also applied for analysis of plasmid DNAs.

Some technical problems interfere with the use of this procedure. Single-stranded nucleic acids form relatively stable hairpins that compete with their hybridization with short oligonucleotides. Incorporation of a universal base (for example, 5-nitroindole) into the terminal positions of immobilized oligonucleotides may solve the problem because it increases the length and thereby the stability of short duplexes (13). In addition, introduction of stabilizing base analogs or replacement of negatively charged phosphodiester groups in oligonucleotides by some neutral or even positively charged groups can significantly increase the stability of duplexes with shorter oligomers relative to that of the hairpins (for references, see ref. 20). The hybridization of long ssRNA or ssDNA with short gel-immobilized
oligonucleotides is a rather slow and low-yield process (19). The hybridization efficiency can be increased by the use of a control electric field to quickly concentrate DNA or RNA in specified chip elements (21).

Successive use of fractionation and sequencing chips could be applied to de novo sequencing or be used in a relatively simple sequence comparison if a homologous fragment is fully sequenced. In this case, a lesser number of longer oligonucleotides (producing more stable duplexes) can be immobilized on the fractionation chips. In the comparative analysis, the sequencing chip will be needed to identify only some changes in the analyzed DNA as compared with the known homologous sequence, rather than to carry out much more complicated de novo sequencing.

The ability of an array of gel pads separated by hydrophobic surfaces to behave as a set of micro test tubes opens up new fields of use for the oligonucleotide chip. Two enzymatic reactions—phosphorylation and ligation of oligonucleotides—were carried out on specified pads of the oligonucleotide chip. The use of more porous and thermostable polyacrylamide (22) or other gels (23) can further expand their application to larger nucleic acid molecules and to other reactions, such as RNA and DNA polymerase reactions and PCR amplification.

The longer a DNA fragment is, the higher is the probability that a certain oligonucleotide will be present in several copies. The repeats interfere with DNA sequencing by hybridization. To increase the sequencing efficiency of the chips, a contiguous stacking hybridization approach has been developed (4,9,12–14). It was demonstrated that one or two 5mers stacked with an immobilized 10mer form stable 15 or 20 bp duplexes upon joint hybridization with DNA (13). This enables one to extend the length of only those immobilized oligonucleotides that are involved in hybridization, and in this way the sequencing efficiency of chips can be increased to become comparable with chips containing longer oligonucleotides. The ligation of two stacked oligonucleotides in a duplex with DNA that had been immobilized on magnetic beads was reported (24). We demonstrate here that phosphorylation and ligation can be carried out site-specifically on chips with gel-immobilized oligonucleotides.

The presence of short oligonucleotide repeats in a DNA fragment can present a serious obstacle to sequencing by hybridization. Measurement of the length of such repeats is an important task, because changes in the repeat length in some genes can be responsible for some genetic diseases (25). Hybridization–ligation ‘walking’ was developed and applied for sequencing tetranucleotide repeats. The microchip-immobilized decamer flanks five ACAT repeats from the 5′-end. The 10mer was hybridized and then successively ligated in four hybridization–ligation ‘walking’ cycles with unlabeled phosphorylated 5mers to establish the 5′-flanking sequence and the repeat length. Alternatively, only one unphosphorylated oligonucleotide 6–50 bases long can be used for hybridization–ligation ‘walking’ if its length is a multiple of the repeat length. Hybridized oligonucleotide should be unphosphorylated, and the enzymatic phosphorylation of the oligonucleotide described above should be added to each hybridization–ligation ‘walking’ cycle to avoid simultaneous co-igation of several copies of the oligonucleotide. Thus, the combined use of contiguous stacking hybridization, phosphorylation and ligation can increase the length of sequenced DNA containing repeats.

It appears that the oligonucleotide microchip can be used as a high-density array to carry out hybridization, fractionation and enzymatic processes on either all chip elements in parallel or in a site-specific way. The potential of this approach lies in the combining of these processes into one procedure for sequence analysis of DNA containing repeats, as well as for sequence comparison of plasmids and cosmids derived from homologous genomic regions without intermediate subcloning.

ACKNOWLEDGEMENTS

We are grateful to Dr V. Shick for helpful discussions. We are indebted to Anne Gemmell, Edward Timofeev, Gennady Yershov and Natalia Mologina for their contribution to this work. We thank Elena Novikova, David Nadziejka and Victor Barsky for their critical comments and editorial assistance. This work was supported by contract 9433352403 with Argonne National Laboratory (USA), by grant #558 from the Russian Human Genome Program, and by the Defense Advanced Research Project Agency under interagency agreement #AO-E428.

REFERENCES