University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Ravi Saraf Publications

Chemical and Biomolecular Research Papers --
Faculty Authors Series

2002

An Approach to Sequence DNA without Tagging

Sanjun Niu

Virginia Tech, Blacksburg, VA

Ravi F. Saraf

University of Nebraska-Lincoln, rsaraf2@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/cbmesaraf>

Part of the [Biomechanics and Biotransport Commons](#)

Niu, Sanjun and Saraf, Ravi F., "An Approach to Sequence DNA without Tagging" (2002). *Ravi Saraf Publications*. 12.

<https://digitalcommons.unl.edu/cbmesaraf/12>

This Article is brought to you for free and open access by the Chemical and Biomolecular Research Papers -- Faculty Authors Series at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Ravi Saraf Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

An approach to sequence DNA without tagging

Sanjun Niu and Ravi F. Saraf *

Department of Chemical Engineering, Virginia Tech, Blacksburg, VA 24060, USA

* Corresponding author.

Abstract

Microarray technology is playing an increasingly important role in biology and medicine and its application to genomics for gene expression analysis has already reached the market with a variety of commercially available instruments. In these combinatorial analysis methods, known probe single-strand DNA (ssDNA) “primers” are attached in clusters of typically $100\ \mu\text{m} \times 100\ \mu\text{m}$ pixels. Each pixel of the array has a slightly different sequence. On exposure to “unknown” target ssDNA, the pixels with the right complementary probe ssDNA sequence convert to double-stranded DNA (dsDNA) by a hybridization reaction. To transduce the conversion of the pixel to dsDNA, the target ssDNA is labeled with a photoluminescent tag during the polymerase chain reaction (PCR) amplification process. Due to the statistical distribution of the tags in the target ssDNA, it becomes significantly difficult to implement these methods as a diagnostic tool in a pathology laboratory. A method to sequence DNA without tagging the molecule is developed. The fabrication process is compatible with current microelectronics and (emerging) soft-material fabrication technologies, allowing the method to be integrable with micro-electromechanical systems (MEMS) and lab-on-a-chip devices. An estimated sensitivity of 10^{-12} g on a $1\ \text{cm}^2$ device area is obtained

1. Introduction

The genetic expression of a specific DNA strand is highly specific, and often unique to an individual. Thus gene profiling is drawing special attention as a powerful tool to make early diagnosis of specific diseases, such as breast cancer, Huntington’s disease and cystic fibrosis. Special traits such as “sweet tooth” may also be determined by gene profiling. Gene-profiling may lead to fast and relatively inexpensive methods to design drugs “personalized” for a particular person or a group of individuals. Apart from diagnostics and drug design, the method can have a broad range of application such as in the analysis of pathogens in the environment and in crime detection. The conventional sequencing technologies resolve the base pair sequence at single-nucleotide level making the analysis effort proportional to the length of the gene or DNA.

Several advances have occurred over the years to accelerate the sequencing process. Notable is the improvement of the conventional sequencing process using gel-electrophoresis. The introduction of parallel, capillary-based electrophoresis systems has greatly increased the rate of diffusion and hence decreased the time during which DNA may be sequenced (Mullikin and McMurray 1999). Although development continues on electrophoresis systems (Hanning *et al.* 2000), the process is linearly proportional to the length of the gene. The significant breakthrough in sequencing technology came in the development of a massively parallel approach to perform many sequences of single-strand DNA (ssDNA) simultaneously. Such a combinatorial chemical analysis and processing approach reduces the amount of time required by the “number of parallel channels.” The method gaining pervasive interest and attention is (in classical terms) a miniaturized version of a reverse dot-blot method. The “smart-blotting”

is a fabricated array of immobilized ssDNA genes of known sequence. The sample ssDNA “dot” may be in solution rather than a gel. The development of such miniaturized concepts has been possible (and triggered by) success in microscale and nanoscale processing technologies primarily spurred by the microelectronics industry (Service 1998). It should also be noted that microelectronics fabrication has had an impact on other biomedical technologies by massive miniaturization of fluid processing. These developments are referred as “lab-on-a-chip” (Burns *et al.* 1998). Thus it is conjectured that in the future a combination of microfluidics (i.e. lab-on-a-chip) and massively parallel probe methods (i.e. biochips) will lead to highly functional and sophisticated hand-held gadgets that may not only out-perform the present devices but they would also be inexpensive and user friendly.

An array of immobilized ssDNA on a substrate is commonly referred to as a “biochip” for DNA sequencing or simply a DNA chip. In biochip methods (first introduced about 1996) a monolayer of specific ssDNA fragments is assembled on an array of pixels ($\sim 1\text{--}100\ \mu\text{m}^2$) (Chee *et al.* 1996). The base sequence in the ssDNA may change from pixel to pixel. These probe fragments act as “chemical tweezers” to pick the specific complementary tagged ssDNA from the sample batch to form double-stranded DNA (dsDNA) (i.e. a hybridization process). The hybridized region is observed by the fluorescent label. The process is highly parallel, very specific and analyses multiple nucleotide sequences (Xu and Shao 2000).

The technology is proving to be highly versatile (Stipp 1997, Wilgenbus and Lichter 1999): for example; Affymetrix Corp. is studying p53 gene malfunction (i.e. mutation) responsible for (breast) cancer; Merck is using the DNA array to study changes in DNA sequencing as the cell begins to rapidly proliferate (thus understand tumor formation); Incyte Pharmaceutical is working on making a disease-specific chip for drug design. Furthermore, the massively parallel, quick, sensitive and accurate biochip methods may boost the Human Genome Project.

Processing is an important aspect of the success of DNA chip technology. Lithographic techniques developed in the microelectronics industry have been key in developing the precise microarrays of ssDNA “blots” (Chee *et al.* 1996). Recently, the processing has further improved the massively parallel processing of over 5,000 spots (i.e. ssDNA probes) on a DNA chip using complete robotic automation (Yoon *et al.* 2000). Apart from the (microelectronics) processing technology, the advancement in optics and electro-optical detection devices is another factor that has driven the biochip technology. The development in confocal microscopes has made the high-resolution read-out of such devices possible (Delpech 2000). Recently, the read-out process has further been automated at high speed and sensitivity by integrating the probe mechanism with photodetectors on the same chip (Vo-Dinh *et al.* 1999). Several other tagging techniques have been developed including, among others, fluorescent, oligonucleotide hairpins (Bonnet *et al.* 1999) and a number of techniques are currently being developed for probing DNA at the single-molecule level (Weiss 1999, Kne-meyer *et al.* 2000). These techniques rely on hybridization of DNA changing the physical properties of the tagged probe. Recently, researchers have reported sequence-selective DNA detection utilizing nanoparticles that polymerize when encountering the target strands (Taton *et al.* 2000).

Apart from the highly parallel approach (i.e. speed), the miniaturization is also an important benefit of biochips. The long-term goal is to make a small portable device that could be used in a doctor’s office to check whether a patient would, for example, (eventually) develop cancer or how fast the body is likely to break down a specific anti-cancer drug. For all these applications, a tool that can perform a DNA fragment analysis of a small sample for several specific genes (all at low concentrations) is highly desirable. The present biochip methods require tagging, the method to make the structure is complex, requiring expensive lithographic methods, and the number of fragments per pixel is not (accurately) known or fixed.

In contrast to present biochip methods, the probing technique discussed below will not require any tagging. Exploiting the change in the corresponding refractive index leads to the contrast between the unhybridized ssDNA probe site and the hybridized double-stranded (dsDNA) probe–target site. Furthermore, compared to present biochip methods the device studied below has a fabrication process based on a (relatively inexpensive and highly parallel) self-assembly approach that does not require expensive lithographic techniques. The simplicity of the processing will allow the device to be integrable with lab-on-a-chip technology.

2. The tagging problem

In most of the present gene profiling methods, the DNA fragments are tagged with a photoluminescent dye. The dye is attached during the polymerase chain reaction (PCR) amplification process. In the PCR process, the DNA is replicated by polymerizing the nucleotides in accordance with the DNA fragment of interest. Since the nucleotides are composed of A, G, T and C bases, if one of the bases is tagged with dye then it will be incorporated in the synthesized DNA during PCR.

In a PCR, let the DNA fragment (typically ~ 25 bases) for replication have n G bases. Let p mole fractions of the G bases in the reaction mixture be tagged with a dye. Assuming no change in reactivity of the base due to tagging, the number fraction of ssDNA synthesized in the PCR with r tags is

$$P(n, r) = \frac{n!}{(n-r)!r!} p^r (1-p)^{(n-r)}. \quad (1)$$

Tagged nucleotide bases are expensive. Thus the mole fraction p is maintained low, usually in the 10% range. Figures 1(a) and (b) show the effect of p on the number fraction of synthesized DNA with no tags (i.e. $r = 0$), one tag (i.e. $r = 1$) and two tags (i.e. $r = 2$) for $n = 5$ and 10. At $p = 0$, $P(n, 0) = 1$ as expected for both $n = 5$ and 10. As expected, $P(n, 0)$ decreases monotonically to 0 as p increases to 1. In contrast, as expected, at low p , $P(n, 1)$ and $P(n, 2)$ increase monotonically as p increases. However, note that, the number fraction is below 40% for the conditions chosen, indicating a high level of inefficiency in the tagging process. Interestingly, $P(n, 1)$ shows a maxima for both $n = 5$ and 10. This is because as p increases there are more DNA fragments with (undesired) multiple tags than with a single tag. $P(n, 2)$ exhibits similar behavior, however for the range shown, the maxima is visible only for $n = 10$, but not for $n = 5$.

Figure 1. Number fraction of DNA synthesized in PCR with n nucleotides that may be tagged. As the p increases, the fraction of untagged nucleotide decreases monotonically. However, the number fraction for $r > 0$ has a maximum.

Figure 2. The η increases with p since more tagged bases are available for insertion. When the number of sites (i.e. n) increases, the probability of DNA with multiple tags also increases. Thus η exhibits a maximum at $n = 10$. The maximum for $n = 5$ occurs at a larger p .

Since, $P(n, 0) > 0$ implies that there are significant numbers of DNA fragments synthesized in the PCR reactor that do not have a tag. We define a tag-efficiency-factor, $\eta = P(n, 1)/(1 + P(n, 0))$ such that $\eta \tau P(n, 1) (< 1)$ for $p = 1$. In an ideal case, if all the ssDNA fragments synthesized have only one tag, $P(n, 1) = 1$, then $\eta = 1$. Figure 2 shows the change in η as p increases. High η is ex-

Figure 3. The background due to multiple tagging increases monotonically as p increases.

Figure 4. The basic element of the DNA chip device is shown. For the study described, the substrate is an Si wafer and the photoluminescent layer is an organic material.

pected as p increases. However, for large values of p , η drops because most of the ssDNA synthesized will have more than one tag.

If we assume that p is small (usually for cost reasons), then the number of ssDNA with more than two tags will be relatively small. For example, for $n = 5$, with p as high as 0.2, $P(5, 0) = 0.328$, $P(5, 1) = 0.401$, $P(5, 2) = 0.205$ and $P(5, 3) = 0.051$, i.e. DNA with three tags is ~ 8 and is four times less than DNA with one and two tags respectively. Thus, $P(n, r)$ for $r > 2$ are ignored. Furthermore, $P(n, 0)$ does not contribute to the photoluminescence (PL) signal. The background due to excess tagging is then defined as the relative amount of ssDNA with two tags instead of one tag. In other words, the background is defined as, $\phi = P(n, 2)/P(n, 1)$. Figure 3 shows the calculated background at $n = 5$ and 10. At $p > 0.2$, there is more ssDNA with two tags than single-tag ssDNA for $n = 10$ and half the amount for $n = 5$.

3. Concept

The elements of the device are shown in Figure 4 (Saraf and Niu 2001). The structure is composed of three layers: the top layer of grafted ssDNA is grafted onto a photoluminescent material and the PL layer (PLL) is deposited on a Si substrate by solution spinning. The substrate may be a fiber-optic probe or a flat wafer. The signal that can be collected is reflection or transmission mode depending on the optical properties of the substrate. For this study, the substrate is a reflector.

Figure 5. PL spectrum from PLL with the top DNA layer being, ssDNA, dsDNA and denatured back to ssDNA. The change in intensity, $I_{PL,ds}/I_{PL,ss} = 2.35 \pm 0.25$.

Upon illumination of the sample with input light of intensity $I_{incident}$, the light intensity on the PLL is

$$I_O = I_{incident} - I_{scattering} - I_{reflection} - I_{absorption} \quad (2)$$

where I_O is the light on the PLL after scattering, reflection and absorption by the top DNA monolayer. The resulting PL from the PLL is then,

$$I_{PL} = KI_O \quad (3)$$

where the constant K is proportional to the total PL efficiency of the PLL and the sensitivity of the detector.

A significant change in the optical properties of DNA is expected as it converts from ssDNA to dsDNA. The contrast depends on the change in refractive index of the top layer as it transforms from ssDNA to dsDNA. This change is expected to be significant because ssDNA is a dielectric material and dsDNA is a semiconductor with a band gap of a few hundred milli-electronvolts (Rakitin *et al.* 2001). Thus the surface refractive index will significantly change as the ssDNA layer transforms to dsDNA.

4. The ssDNA to dsDNA contrast

The target ssDNA layers in figure 4 are attached to the surface using the standard method described in the literature (Brockman *et al.* 1999). In this study, the ssDNA target is attached covalently to the amine functionalized PLL surface. The sequence of the immobilized ssDNA, called the probe, is CAA-AAT-AGA-CGC-AAC-GAA-AAC and its complement target ssDNA is ATA-GTT-TTC-GTT-GCG-TAA-GCG-TCA-ATT. The probe ssDNA has a thio-group at the 5'-end. The surface concentration of grafted ssDNA, as measured by PL, is discussed later in section 5. From the control experiments using tagged probe ssDNA, the per cent immobilization is within 8%. However, we note in passing, since the method measures the signal before and after the hybridization step for each pixel, that the reproducibility of immobilization is not critical. A tolerance of up to 15% should be sufficient for the excellent signal-to-noise ratio (SNR) obtained (as discussed later with respect to figure 5). The grafting is performed at the 5'-end. Note also that the first three bases, CAA,

are not hybridized. They provide flexibility to the structure at the DNA/substrate interface. The provision of such a spacer is important to achieve fast hybridization kinetics. The concentration of target ssDNA is $6.6 \times 10^{-4} \text{ nmol } \mu\text{l}^{-1}$ or $5.5 \text{ ng } \mu\text{l}^{-1}$. The incubating time for grafting is 5 h at 50°C and the hybridization process is 10 h at 40°C .

The change in PL from PLL as the top DNA layer in figure 4 changes from ssDNA (probe) to dsDNA (probe-target) is measured in a spectrophotometer at an excitation wavelength of 265 nm. Figure 5 shows the photoluminescent intensity, I_{PL} , from PLL for ssDNA and after hybridization to dsDNA. The contrast is highly amplified compared to the expected difference due to simple absorption. The photoluminescent intensity for ssDNA, $I_{PL,ss}$, is lower than $I_{PL,ds}$ for a dsDNA top layer. Thus the light leaking into the PLL is larger when the top layer is dsDNA. Although the optical density is consistent with this result, the major change occurs because of changes in reflectivity and scattering.

5. Sensitivity estimate

We developed an assay using a photoluminescent dye o-phthalaldehyde (OPA) that readily attaches selectively to amine groups. By calibrating the PL from OPA solution, it is estimated that the number of amine groups is $2.5 \times 10^{10} \text{ moieties cm}^{-2}$. The reproducibility of the surface modification is within 10%. Assuming all the amine species attach to an ssDNA during the immobilization step, the maximum number of ssDNA immobilized on a surface is $2.5 \times 10^{10} \text{ molecules cm}^{-2}$. This is a conservative guess for the estimated sensitivity because the actual number of ssDNA would be less due to steric hindrances and, possibly, the electrostatic effect. Thus, for the present DNA chain length, the measured signal (i.e. figure 5) is due to at most $\sim 2.4 \times 10^{-10} \text{ g cm}^{-2}$ of 20 base long sample ssDNA at an SNR of better than 100:1. The effective sensitivity is further improved by two orders of magnitude by relaxing the SNR to 10:1 and increasing the detector sample time from 50 ms to 1 s.

6. Sensitivity to single-nucleotide polymorphism (SNP)

We define the change in PL from the PLL as the top layer changes from ssDNA to dsDNA as, $R(x) = I_{PL,ss}/I_{PL,ds}$, where x is the number of mismatched nucleotides between the probe ssDNA and the target ssDNA, i.e. x is the number of SNPs. Thus for a perfect match, $x = 0$ and for one SNP, $x = 1$.

Figure 6 indicates the possibility of measuring one SNP. As the hybridization temperature increases, the binding for both a perfect match and one mismatch decreases. At lower hybridization temperatures, $R(0)$ and $R(1)$ are > 1 indicating that both the perfectly matched sample and the one SNP sample bind to the probe ssDNA. At around $\sim 48^\circ\text{C}$, the one SNP sample does not bind, i.e. $R(1) = 1$, however, for a perfect match, $R(0) > 1$. Although the difference is only 9.3% which is barely above the error bar, an increase in the contrast may improve the measurement. Currently a change in optics with a higher numerical aperture lens to collect the PL at higher solid angle is being pursued.

Figure 6. Behavior of hybridization temperature on DNA binding exhibits $< 2^{\circ}\text{C}$ sensitivity. $R(x) > 1$ indicating that the PL increases upon DNA hybridization as seen in figure 5.

7. Conclusion

In microarray technology for DNA gene profiling, transduction is performed by measuring the change as the known probe ssDNA, immobilized on a specific pixel site, hybridizes with the unknown target ssDNA. The target ssDNA is usually tagged during an amplification process in a PCR. The statistical error due to the tagging during the PCR can be significant enough to make most of the array technologies difficult to implement as diagnostic tools in hospitals. A new optical method is described to perform gene profiling without tagging the target. The results indicate a good contrast with SNR better than 500:1. The device is a simple three-layer structure that can be integrated with a lab-on-a-chip device as a detection system.

Acknowledgments

The authors would like to acknowledge Carilion Biomedical Institute, Roanoke, VA and Optical Science and Engineering Research Center, Virginia Tech for the financial support.

References

- Bonnet, G., Tyagi, S., Libchaber, A., and Kramer, F. R., 1999. *Proc. Natl Acad. Sci. USA* **96**: 6171–6176
- Brockman, J. M., Frutos, A. G., and Corn, R. M., 1999. A multistep chemical modification procedure to create DNA arrays on gold surfaces for the study of protein-DNA interactions with surface plasmon resonance imaging *J. Am. Chem. Soc.* **121**: 8044–8051
- Burns, M. A., *et al.*, 1998. An integrated nanoliter DNA analysis device *Science* **282**: 484–487
- Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, Z. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., and Fodor, S. P. A., 1996. Accessing genetic information with high-density DNA arrays *Science* **274**: 610–614
- Delpech, M., 2000. DNA chips *Ann. Biol. Clin. (Paris)* **58**: 29–38
- Hanning, A., Lindberg, P., Westberg, J., and Roeraade, J., 2000. Laser-induced fluorescence detection by liquid core waveguiding applied to DNA sequencing by capillary electrophoresis *Anal. Chem.* **72**: 3423–3430
- Knemeyer, J. P., Marmé, N., and Sauer, M., 2000. Probes for detection of specific DNA sequences at single-molecule level *Anal. Chem.* **72**: 3717–3724
- Mullikin, J. C., and McMurray, A. A., 1999. Sequencing the genome, fast *Science* **283**: 1867–1868
- Rakitin, A., Aich, P., Papadopoulos, C., Kobzar, Y., Vedeneev, A. S., Lee, J. S., and Xu, J. M., 2001. Metallic conduction through engineered DNA: nanoelectronics building blocks *Phys. Rev. Lett.* **86**: 3607–3673
- Saraf, R. F., and Niu, S., 2001. *Patent Pending*
- Service, R. F., 1998. Microchip arrays put DNA on the spot *Science* **282**: 396–433
- Stipp, D., 1997. Gene Chip Breakthrough *Fortune* **March 31**: 56–73
- Taton, T. A., Mirkin, C. A., and Letsinger, R. L., 2000. Scanometric DNA array detection with nanoparticle probes *Science* **289**: 5485
- Vo-Dinh, T., *et al.*, 1999. DNA biochip using a phototransistor integrated circuit *Anal. Chem.* **71**: 358–363
- Weiss, S., 1999. Fluorescence spectroscopy of single biomolecules *Science* **283**: 1676–1683
- Wilgenbus, K. K., and Lichter, P., 1999. DNA chip technology ante portas *J. Mol. Med.-JMM* **77**: 761–768
- Xu, B. S., and Shao, J. Z., 2000. Advancement of several new types of biochip *Prog. BioChem. BioPhys.* **27**: 251–254
- Yoon, S. H., Choi, J. G., and Lee, S. Y., 2000. Development of DNA chip microarrayer *J. Microbiol. BioTech.* **2**: 21–26