Automated parallel DNA sequencing on multiple channel microchips


Molecular Dynamics/Amersham Pharmacia Biotech, Sunnyvale, CA 94086

Communicated by I. Robert Lehman, Stanford University School of Medicine, Stanford, CA, March 14, 2000 (received for review January 4, 2000)

We report automated DNA sequencing in 16-channel microchips. A microchip prefilled with sieving matrix is aligned on a heating plate affixed to a movable platform. Samples are loaded into sample reservoirs by using an eight-tip pipetting device, and the chip is docked with an array of electrodes in the focal plane of a four-color scanning detection system. Under computer control, high voltage is applied to the appropriate reservoirs in a programmed sequence that injects and separates the DNA samples. An integrated four-color confocal fluorescent detector automatically scans all 16 channels. The system routinely yields more than 450 bases in 15 min in all 16 channels. In the best case using an automated base-calling program, 543 bases have been called at an accuracy of >99%. Separations, including automated chip loading and sample injection, normally are completed in less than 18 min. The advantages of DNA sequencing on capillary electrophoresis chips include uniform signal intensity and tolerance of high DNA template concentration. To understand the fundamentals of these unique features we developed a theoretical treatment of cross-channel chip injection that we call the differential concentration effect. We present experimental evidence consistent with the predictions of the theory.

The initiatives to complete the sequence of the human genome by 2003 (1) and the draft sequence as early as spring 2000 (2) demand cost-effective high-throughput, high-performance sequencing technologies. DNA sequencing separations traditionally have been performed on slab gels (3). Recently, capillary array electrophoresis (CAE) (4–9) has been demonstrated to be a high-speed, high-throughput method for DNA sequencing. CAE instrumentation is being adapted in leading genomics centers to complete the human genome sequence. The throughput of a CAE system is directly proportional to the number of separation capillaries in the instrument. However, as the number of capillaries increases, it becomes more challenging to control sample injection and detect signals from all of the capillaries.

An even more advanced technology for high-throughput DNA analysis is CAE on microchips. Because photolithographic techniques are used to make CAE microchips, additional capillaries can be readily added. Several labs have used microchips to analyze oligonucleotides and RNA, and to genotype and sequence DNA. The analyses are rapid because of the short separation channel and efficient because of the injection of narrow sample bands.

Manz and coworkers (10) introduced photolithographic technologies to microfabricate electrophoretic separation channels in 1992. Microfabricated capillary electrophoresis (CE) devices have been used to separate fluorescent dyes (11, 12), fluorescently labeled amino acids (13–16), DNA restriction fragments (17, 18), PCR products (17, 19), short oligonucleotides (20), short tandem repeats (21), and DNA sequencing fragments (22–25). Separations on CE chips are extremely rapid and normally are complete in seconds to minutes, whereas ultrahigh speed separations can be finished in milliseconds (26) or submilliseconds (27). The highest speed sequencing results (22) used a 7-cm-long electrophoresis channel. Separations of 500 bases in four colors were complete in 20 min with an accuracy of greater than 99%. Although this demonstrated the fundamental feasibility of four-color DNA sequencing on microfabricated devices, high-throughput, multichannel DNA sequencing was not achieved because only a single channel was used.

To increase throughput, CAE channels have been microfabricated on microchips and successfully used for DNA fragment size analysis (23, 28–30). Channels on these devices have right-angle turns that work well for fragment sizing but degrade sequencing separations. In another design (31), 96 straight channels without turns were fabricated on a 10-cm-diameter CAE chip extending radially from the center of the chip where a common anode reservoir is located. The sample, waste, and cathode reservoirs were arranged around the circumference of the chip. This design, combined with a rotary scanning detection system, provides unique features: (i) chip space is effectively used, (ii) the detector scans perpendicularly across all separation channels, and (iii) all channels are designed to be identical to facilitate uniform sample injections and separations. The disadvantage of this design is that the separation channel length is limited to less than half of the chip diameter: only 3.3-cm effective separation lengths were obtained on a 10-cm-diameter device. Channels of this dimension work well for separations of certain restriction fragments and genotyping samples (31), but it is challenging to achieve long read lengths in sequencing separations with such short channels.

To realize the benefit of high-speed, high-throughput separations on CAE microchips, automated microchip preparation, including matrix replacement and sample loading, will be required. In this work, we report DNA sequencing on a 16-channel CAE chip with automated sample loading using a robotic instrument. The CAE chip is fabricated on a 10-cm-diameter wafer (see Fig. 1) with effective separation lengths (from injector to the detection region) ranging from 7 cm (channels in the middle) to 7.6 cm (channels on both sides). All channels have the same length from cathode to anode. Samples are automatically transferred from a 96-well plate to a 16-channel chip by using an eight-tip pipettor. The chip then is automatically moved, aligned, and focused to a detector. Voltages automatically are applied to various reservoirs by using an electrode array board (30). Detection is provided by a four-color scanning confocal laser-induced fluorescence detector with a scan rate of 9.4 Hz. Sequencing separations including chip preparation and sample loading are completed in about 18 min and bases are called by using a MegaBACE base-calling program. An accuracy >99% was achieved for up to 543 bases.

Experimental Procedures

CAE Chip Design. Fig. 1 presents the CAE chip design used in this experiment. The 16-channel chip is a combination of two...
Chip Preparation. Linear polyacrylamide (LPA) (4% wt/vol), prepared according to the procedure described in ref. 22, was forced under pressure (200 psi) into all channels of the chip through two common anode reservoirs until excess sieving matrix reached all reservoirs. The excess sieving matrix in reservoirs was washed out with water. The anode, cathode, and waste reservoirs then were filled with diluted MegaBACE LPA buffer solution (Amersham Pharmacia Biotech). The entire process took less than 5 min. The same procedure was used to regenerate the chip. Chips have been run up to 50 times without degradation.

Instrumentation. The detailed instrumental automation will be published elsewhere. Briefly, the instrumentation automatically warms up the chip with a temperature-controlled aluminum plate, loads samples from microtiter plates into the 16-channel CAE chip by using an eight-channel pipettor, moves the chip to the detection position, electrophoretically separates the samples, detects the separated fluorescent fragments, and exports the data for analysis.

Four-color detection was performed on a laser-induced confocal scanning detector. A 488-nm line from a multiline 40 mW Ar+ laser (Uniphase, San Jose, CA) was passed through a laser line filter and directed by a primary beam splitter to a galvonometer scan mirror. The excitation beam was reflected through a telecentric scan lens (33) and focused onto the chip channels. The telecentric scan lens worked in consort with the chromatic correction lens in the emission path.

The laser excitation was delivered to the chip orthogonal to the chip surface. The scan lens collected fluorescent light from the DNA samples and returned it to the galvonometer mirror, which descanned the beam and directed it through the primary beam splitter. The emission beam then was directed through a 500-nm long-pass laser-blocking filter and a combined chromatic correction lens and achromat that focused the beam onto the confocal pinhole. The pinhole spatially filtered the emission light and fed an achromat that collimated the emission beam for separation by secondary and tertiary beam splitters. These dichroic beam splitters divided the emission beam into four spectral regions (500–540, 540–570, 570–595 and >595 nm). The fluorescence signals were further filtered by using 520/20, 555/20, 585/20 bandpass filters and a 610 longpass filter before entering four photomultiplier tubes (Hamamatsu, Bridgewater, NJ). The limit of detection of the system is 40 pM of fluorescein.

Data Acquisition and Processing. Four 16-bit analog-to-digital converter boards (ComputerBoards, Middleboro, MA) were used in the detection system, one for each photomultiplier, at a 50-kHz acquisition rate. Every four data points were averaged to create one picture element (pixel). Each pixel mapped to an area of 10 μm in the galvanometer scan dimension; 1,000 pixels were acquired during each forward scan of 16 channels. The overall galvanometer scan frequency was 9.4 Hz. The galvanometer-based detector created four two-dimensional images, one for each photomultiplier. One dimension represents the galvanometer scan dimension and the other dimension represents the time dimension (refer to Fig. 4).

A software program BBUTIL (Molecular Dynamics) was used to convert the two-dimensional images into a series of electropherograms corresponding to each of the CAE separation channels. The center pixels of the separation channels were automatically detected by BBUTIL, and 15 pixels (the center pixel plus seven pixels on each side) along the scan dimension were weight-averaged to generate a data point of the electropherogram. A “trapezoidal” averaging method was used to place more weight on the center pixels versus the side pixels.

The electropherograms were saved in either a text file or a binary (.ESD) format compatible with Molecular Dynamics.
MegaBACE analysis software. The raw data in .ESD format were processed and base-called by the MegaBACE base-calling program, CIMARRON PHAT. The final sequence was exported in a corresponding separation field strength was calculated for waste reservoirs, while grounding the cathode reservoir. The sample and anode reservoirs were connected by applying 2,000 V to the anode reservoir, 140 V to sample and waste reservoirs, typically 10 V, respectively to the waste and cathode reservoirs, typically 227 V/cm. Barbara, CA).

Electrophoresis Methods. The platinum wire electrodes were arranged according to the locations of chip reservoirs and affixed to a fiberglass board (30). Anode, cathode, sample, and waste reservoirs were grouped together and connected to four separate programmable high-voltage power supplies (EMCO, Sutter Creek, CA) through high-voltage relays (Kilovac, Santa Barbara, CA).

Sample injection was performed by applying voltages of 50 and 10 V, respectively to the waste and cathode reservoirs, typically for 60 s, while the sample and anode reservoirs were grounded. Separations were carried out immediately after sample injection by applying 2,000 V to the anode reservoir, 140 V to sample and waste reservoirs, while grounding the cathode reservoir. The corresponding separation field strength was 227 V/cm. Energy transfer-dye-labeled primer M13 MegaBACE standard sequencing samples (1×, Amersham Pharmacia Biotech) were exclusively used in this experiment, each separation consumed 1-μl sample.

Results

Effect of Injection Time on Separations. After determining appropriate electrical biasing conditions (12), we investigated the length of time required for sample injection. Fig. 2 shows five electropherograms using injection times of 10, 20, 30, 50, and 70 s, all on the same scale. At an injection time of 10 s, none of the peaks achieved steady-state injection. At an injection time of 20 s, fragments of fewer than 270 bases reached their steady-state levels; however, longer fragment peaks were not observed in these electropherograms because they did not reach the separation channel before the start of separation. After 30 s of injection, fragments of 550 bases and smaller obtained steady-state levels. All sequencing fragments reached their steady-state levels after 50-s injection and were unchanged at 70 s of injection.

The separations in Fig. 2 were performed under low field strength and room temperature to magnify the injection time effects. As a result, the separations were slow compared with routine separations.

Effects of Temperature on Separations. Elevated temperature is important to achieve long read lengths and accurate DNA sequencing results using CE (34–36) and on CE chips (22). Increased column temperature not only reduced compressions and improved resolution, but also accelerated separation (34–36).

Investigation of the temperature effect on sequencing in 16-channel CAE chips from 20°C to 60°C indicated that the resolution for fragments longer than 500 bases decreased slowly upon increasing temperature (data not shown). This observation is consistent with the CE sequencing literature (37). We performed most of our experiments at 50°C, which was a compromise between resolving short and long DNA fragments. At this temperature, compressions were reduced because of the increased denaturation of the DNA. The electrophoretic mobility doubled as the temperature was elevated from ambient temperature to 50°C.

Tolerance of Template. Fig. 3a presents the effect of template concentration on sample injection. These separations were performed with a 50-s injection time under the same experimental conditions as Fig. 2. The template concentrations in regular MegaBACE sequencing standards contained 50 ng/μl DNA template. As the template concentration was increased to 40 ng/μl, fragments of >600 bases were not observed. More fragments disappeared when the template concentration increased to 70 ng/μl. To recover these missing bases, a longer injection time is required, as shown in Fig. 3b for a sample containing 40 ng/μl template.

Four-Color DNA Sequencing on 16-Channel CAE Chips. After optimizing injection and separation conditions, we tested the CAE chip for four-color DNA sequencing in 16 channels. Fig. 4 presents a typical raw image from the photomultiplier that corresponds to T-terminated fragments. The 16 channels are clearly shown in...
the scan dimension as is the separation in the time dimension. The last peak at the end of the image (15 min) corresponds to 440 bases for the first (from the bottom) channel and 490 bases for the eighth channel. High-quality separations are achieved in all 16 channels more than 90% of the time.

Fig. 5 shows the results of using base-calling software (Cimarron, Salt Lake City, UT, optimized for MegaBACE) to process the data from the fourth channel in Fig. 4. The raw data in .E3D format were automatically analyzed by the program. Five errors were found in the base-called sequence up to 543 bases; this corresponds to greater than 99% accuracy for a 16-min separation. Base calling all 16 lanes produced an average read length of 457 ± 35 bases for an accuracy of 99% and 512 ± 44 bases for an accuracy of 98% (Table 1).

**Discussion**

**Considerations of the Chip Design.** We observed four main design constraints to make a practical multiple-channel CAE chip for automated high-throughput DNA sequencing: (i) To interface

<table>
<thead>
<tr>
<th>Channel no.</th>
<th>Read length at 99% accuracy</th>
<th>Read length at 98% accuracy</th>
<th>Migration time of 550-base fragments, min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>472</td>
<td>522</td>
<td>17.3</td>
</tr>
<tr>
<td>2</td>
<td>518</td>
<td>568</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>446</td>
<td>501</td>
<td>16.5</td>
</tr>
<tr>
<td>4</td>
<td>543</td>
<td>608</td>
<td>16.1</td>
</tr>
<tr>
<td>5</td>
<td>456</td>
<td>576</td>
<td>15.9</td>
</tr>
<tr>
<td>6</td>
<td>421</td>
<td>500</td>
<td>15.9</td>
</tr>
<tr>
<td>7</td>
<td>443</td>
<td>479</td>
<td>15.9</td>
</tr>
<tr>
<td>8</td>
<td>427</td>
<td>458</td>
<td>15.7</td>
</tr>
<tr>
<td>9</td>
<td>442</td>
<td>490</td>
<td>16.1</td>
</tr>
<tr>
<td>10</td>
<td>421</td>
<td>478</td>
<td>16.1</td>
</tr>
<tr>
<td>11</td>
<td>459</td>
<td>509</td>
<td>16.2</td>
</tr>
<tr>
<td>12</td>
<td>433</td>
<td>450</td>
<td>16.2</td>
</tr>
<tr>
<td>13</td>
<td>475</td>
<td>517</td>
<td>16.5</td>
</tr>
<tr>
<td>14</td>
<td>433</td>
<td>513</td>
<td>17</td>
</tr>
<tr>
<td>15</td>
<td>488</td>
<td>552</td>
<td>17.6</td>
</tr>
<tr>
<td>16</td>
<td>439</td>
<td>478</td>
<td>17.6</td>
</tr>
</tbody>
</table>

Average ± Std: 457 ± 35, 512 ± 44, 16.5 ± 0.6
between macroworld samples in a 96-well microtiter plate and the microworld of a chip using an eight-tip pipettor, we located cathode, sample, and waste reservoir at spacings compatible with an eight-tip pipettor. (ii) All separation channels should be straight because curved channels reduce separation resolution. (iii) Channels in the detection region must be within the scanning distance of the detector, 10 mm for our detector. (iv) All channels should have the same total length from anode to cathode to facilitate equal electrical field strengths.

The design presented in Fig. 1 meets these constraints. Cathode, sample, and waste reservoirs are arranged in a line and evenly spaced at 4.5-mm intervals. This arrangement allows the use of an eight-channel pipettor with two cycles of the pipettor loading all 16 reservoirs. The radial design enables straight separation channels to be formed from the injectors to the detection region. In the detection region, channels are separated by 600 μm (center to center); 16 channels are within the 10-mm scan region. This channel geometry resulted in different effective separation distances (from the injector to the scan line). These distances varied from 7.0 to 7.6 cm, respectively, from the middle to the side channels, which is why peaks in the middle channels appeared earlier (see Fig. 4) than peaks in side channels. The total channel lengths for all channels are the same, 10.7 cm from injector to anode reservoir. Because the merged channels after the scan line are wider than single channels, the total electrically equivalent channel length is less than 10.7 cm. For 50-μm deep channels used in these experiments, the equivalent channel length is ca. 8.8 cm. No significant read-length decrease was observed from side to middle channels (see Table 1).

Differential Concentration Effect and Steady-State Injection. One striking difference between separations in CAE chips and CAE in pulled capillaries is the improved balancing between the signal intensities of short and long sequencing fragments (22, 24). This is caused by a unique characteristic of a chip injector for DNA fragment separations.

As schematically shown in Fig. 6, when a positive voltage is applied to the waste, DNA fragments migrate from the sample reservoir, across A1 (the water/gel interface, assumed to be perpendicular to the channel axis) to the side channel entrance, to the injector (the cross section area), then across A2 (a plane immediately after the injector and parallel with A1) to the side channel exit, and finally to the waste reservoir. For simplicity we have represented the twin T as a cross channel. The flux across A1 and A2 can be expressed as,

\[ J_{I,A1} = [\text{DNA}]^{\mu g} \mu^g E^8 \tag{1} \]

and

\[ J_{I,A2} = [\text{DNA}]^{\mu gb} \mu^g E^8 \tag{2} \]

where [DNA] and \( \mu \) represent the concentration and mobility of a DNA fragment, \( E \) the electric field strength, subscripts \( w \) and \( g \) represent water and gel phases and subscripts \( f \) the lengths of the DNA fragments in nucleotides.

A steady injection state of DNA fragments, defined as the state in which DNA fragment concentrations in the injector don’t change with time, can be reached as long as the sample in the sample reservoir is not depleted. That is, \( J_{I,A1} = J_{I,A2} \). At steady state,

\[ [\text{DNA}]^g = \frac{\mu^g E^w}{\mu^g E^8} [\text{DNA}]^w. \tag{3} \]

The mobility of a DNA fragment is greater in aqueous solution than in gel and at the interface \( (A1) \), \( E^w \) is larger than \( E^8 \) if the DNA sample is pure. Therefore, the fragment concentration in the gel is higher than in the sample, and fragments are concentrated in the gel phase.

Replacing the subscript \( I \) with \( L \) for a large fragment and \( S \) for a small fragment, and assuming a constant mobility for both large and small fragments in aqueous solution, Eq. 4 is generated by dividing [DNA] \( ^{L} \) with [DNA] \( ^{S} \).

\[ \frac{[\text{DNA}]^L}{[\text{DNA}]^S} = \frac{(\mu^L)^w}{(\mu^L)^8} [\text{DNA}]^{w}_{L} / [\text{DNA}]^{w}_{S}. \tag{4} \]

This equation states that the relative concentration of large fragments to small fragments is increased by a factor of \( \mu^L/\mu^S \) after they enter the gel phase. We call this the differential concentration effect. When the separation voltage is engaged, differentially concentrated DNA fragments at the injection site are injected into the separation channel.

Because \( \mu^L/\mu^S \) is greater than one, the differential concentration effect is biased toward larger DNA fragments. In contrast, electrophoretic injection is biased toward short fragments (38, 39). In DNA sequencing samples, small fragments usually have higher concentrations than large ones. The differential concentration effect associated with CE chips compensates for these concentration differences and explains why uniform signal intensity profiles usually are observed (22, 24).

It follows from the above that the times for DNA fragments to reach a steady state at the injection site also will depend on the size of the fragments. A general equation can be used for calculation of these times, that is,

\[ t_I = \frac{d}{v^g} = \frac{d}{\mu^g E^8}, \]

where \( d \) represents the distance between \( A1 \) and \( A2 \), and \( v^g \) and \( \mu^g \) respectively are the electrophoretic velocity and mobility of the DNA fragments in the gel.

The injection time data obtained in Fig. 2 match the values predicted by Eq. 5 very well. The linear velocities of primer, 267-base, 550-base, and termination peaks were calculated to be 208, 89.0, 52.5, and 35.7 μm/s, respectively based on migration times shown in Fig. 3. Referring to Fig. 6, \( d \) was measured on an actual chip to be 1.8 mm. Based on Eq. 5, the steady-state injection times were calculated to be 8.7, 20, 34, and 50 s, respectively for the mentioned peaks, which correlates well with the experimentally measured numbers, 10, 20, 30, and 50 s.

Removal of template molecules has been reported to be essential to achieve high-quality separations (37, 38, 40). This issue became less important when a twin-T injector was used on a CE chip because template molecules were automatically
removed (22). Our further investigations show that DNA template affects sample injection and subsequently the sequencing separation, as indicated in Fig. 3e. We explain these results as follows. When template (a long polymer) molecules enter the entrance side channel filled with gel, they are entangled with LPA molecules. The entangled DNA-LPA polymer mixture is equivalent to a sieving matrix of increased concentration or molecular weight of LPA. The entangled DNA also may hybridize with the sequencing fragments. As a result, the electrophoretic mobility of the DNA fragments behind the template DNA is reduced. Consequently, it takes longer for large fragments to reach a steady state at the injection site. An alternative explanation is that the template simply builds up at the entrance of the injection channel, resulting in a high resistance region. This causes more of the potential to be dropped in this region and the injection is slowed down because of a decreased field strength. These hypotheses suggest that we should be able to recover the missing bases by lengthening the injection time. Fig. 3b demonstrates experimental verification of this prediction. The time for the biased reformation fragments to reach steady state was almost doubled as the template concentration increased from 20 to 40 ng/μl, implying a 50% reduction of mobility during injection. This observation supports the hypothesis.

Conclusions. We have demonstrated parallel DNA sequencing on multiple channel chips that were automatically loaded and run. Separations in less than 15 min normally resolved more than 450 bases at 99% accuracy with all 16 channels operational 90% of the time. This is a milestone in the development of automated microchip-based DNA sequencers.

Increased read lengths depend on the improvements in the base-calling software, as well as the separations. The CMARRON base-calling program was developed for capillary electrophoresis in long capillaries and is not optimized for data obtained on chips. CAE chip separation data have a different migration peak width and peak spacing profile than CAE in long capillaries. Further optimization of the base-calling program for chip-based DNA sequencing is in progress to improve the base-calling accuracy and read lengths.

A project to sequence on 15-cm-diameter chips recently has been initiated in our laboratory. We expect to fabricate 48 or 96 channels on these chips. With the increased separation distance and optimized software, an average read length of 600 bases with a run-to-run time of less than 25 min should be attainable.

We gratefully acknowledge sponsorship from the National Institutes of Health through Grants R01HG01775–03 and R34HG02980–01 and the National Institute of Standards and Technology through Grant 70NANB5H131.