Annu. Rev. Biomed. Eng. 2001. 3:195–223

NEW DNA SEQUENCING METHODS

Andre Marziali[1] and Mark Akeson[2]
[1]*Department of Physics and Astronomy, 6224 Agricultural Road, University of British Columbia, Vancouver, BC, Canada, V6T-1Z1; e-mail: andre@physics.ubc.ca*
[2]*Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064; e-mail: makeson@chemistry.ucsc.edu*

Key Words electrophoresis, capillary, genomics, microfabrication, ion channels

■ **Abstract** The Human Genome Project and other major genomic sequencing projects have pushed the development of sequencing technology. In the past six years alone, instrument throughput has increased 15-fold. New technologies are now on the horizon that could yield massive increases in our capacity for de novo DNA sequencing. This review presents a summary of state-of-the-art technologies for genomic sequencing and describes technologies that may be candidates for the next generation of DNA sequencing instruments.

CONTENTS

INTRODUCTION

Developed in 1977, the Sanger method (1) of sequencing DNA has revolutionized the biological sciences. In particular, development of an automated fluorescent sequencer based on the Sanger method (2, 3) led the US Department of Energy and the National Institutes of Health to consider sequencing the entire human genome, and in 1990, the Human Genome Project (4, 5) began. Its ambitious goals

have spurred researchers to improve DNA sequencing technology. This resulted in such rapid increases in instrument throughput that most of the human genome was sequenced in only the last two years of the project.

This review presents the most recent developments in DNA sequencing technology and describes new strategies that may lead to the next generation of DNA sequencers. Because of space limitations and the broad nature of this topic, we focus primarily on de novo genomic sequencing and on technical parameters, such as sequencing speed, read length, and base-call precision. However, several highly parallel sequencing strategies have emerged in the past few years that are currently not suitable for genomic sequencing. We include these to provide a more comprehensive picture of the field. Discussion of technologies in current use focuses on the most widely adopted state-of-the-art instrumentation; discussion of newer, developing technologies focuses on theoretical underpinnings. Where possible, the reader is directed to detailed topical reviews.

ELECTROPHORETIC METHODS

Slab Gel Electrophoresis

Acrylamide slab gel electrophoresis has, until recently, been the dominant technology for de novo sequencing of DNA fragments. Initially applied with the use of radioactive (^{32}P) labels and autoradiograms, this technology saw a major advance with the release of the Applied Biosystems (ABI) (Foster City, CA) PRISM 373, the first widely adopted fluorescent detection–based slab gel sequencer. In this instrument, fluorescently labeled DNA fragments from a Sanger sequencing reaction are loaded into wells at the top of a thin (250 μm), cross-linked acrylamide gel sandwiched between two plates of glass. An electric potential is applied between the top and bottom of the gel, causing the fragments to migrate through the gel and separate owing to length-dependent viscous drag. A scanning fluorescence detection system detects fragment bands as they drift past a region at the bottom of the gel. This optical detection system is typically an argon-ion laser focused on the gel through a confocal objective that also collects the fluorescence and directs it to photomultiplier tubes (PMT) or to a charge coupled device (CCD) array. Reconstruction of the original nucleotide sequence of the sample from the fluorescent data is performed by base-calling software that can also assign confidence values (6, 7) to the called nucleotides, making eventual fragment assembly faster and more reliable.

The ABI 373 went through several throughput and read-length upgrades before being replaced by the still popular ABI PRISM 377, which in its current form uses four-color fluorescent dye chemistry coupled with CCD imaging and can process up to 96 samples per 9- to 11-h run. Read lengths of 650–750 bases can be expected from this instrument, which is still a widely used sequencing platform despite suffering from the common drawbacks of slab gel instruments: gel casting, gel loading, and lane tracking. Numerous other slab gel instruments

have been produced either as prototypes or commercially, and these are covered in recent reviews (8, 9). Of these, a few devices deserve mention because, despite the advent of capillary array electrophoresis, slab gels continue to be the preferred technology for very long read length or low-throughput sequencing.

LI-COR Inc (Lincoln, NE) and NEN Life Science Products (Boston, MA) produce the global IR^2 4200 automated DNA sequencer based on a two-color, near-infrared (IR) dye chemistry (IRDye 800 Terminators). Presumably as a result of low background fluorescence in the near-IR, this instrument is capable of read lengths in excess of 1000 bases per sample at 99% accuracy and has therefore found a market in applications requiring extremely long reads. Throughput from these 64-lane instruments is enhanced by the ability to simultaneously sequence both directions on a template. A major feature of the Global IR^2 system is novel software that allows highly automated instrument operation and remote access through any internet connection.

Another development recently made available through MJ Research Inc (Waltham, MA) is that of horizontal ultrathin gel electrophoresis (10). Though still utilizing a slab gel system, the BaseStation uses a 75-μm thick polyacrylamide gel to improve heat dissipation in the gel, thus allowing higher electric fields to be utilized and resulting in faster run times. This, coupled with robotic gel loading, results in an instrument that appears to offer two of the main advantages of capillary electrophoresis with the added benefit of long reads. The instrument is capable of reads ranging from about 500 bases in 1.5 h to over 1000 bases in 6 h. Detection is performed through a four-color photomultiplier system capable of detecting as little as 10 attomol of labeled DNA. Up to 100 samples can be run on each gel.

With the exception of the MJ BaseStation, gel loading is a significant problem for users of slab gel instruments. In an attempt to alleviate this, some novel sample loading schemes have been developed, including membrane-mediated loading (11) (MWG Biotech, Inc, High Point, NC; Sigma-Aldrich, St. Louis, MO). In these schemes, the samples to be sequenced are spotted on the teeth of a thin, porous membrane that is then inserted between the gel plates to be brought into direct contact with the edge of the polymerized gel. This method allows the membrane to be automatically loaded from a variety of liquid-transfer devices or liquid-spotting instruments, and also it allows higher lane density and lower likelihood of sample cross-talk, which can occur from well leakage in conventional slab gel loading.

Capillary Electrophoresis

The persistent drawbacks of slab gel electrophoresis and the desire for faster sequencing runs and higher throughput led to the development of capillary array electrophoresis (CAE) (12). Electrophoresis works in a way similar to slab gels, except that each capillary contains a single sample and therefore tracking problems are eliminated. Furthermore, the high surface-to-volume ratio of a capillary allows more rapid heat dissipation than is possible in slab gels, thus allowing higher operating voltages and faster run times.

The Molecular Dynamics (Sunnyvale, CA—a division of Amersham Pharmacia Biotech) MegaBACE 1000 was the first commercial CAE instrument introduced, followed closely by the Applied Biosystems PRISM 3700. These two instruments continue to represent the state-of-the-art in commercially available DNA sequencing instrumentation, and because of relatively similar performance, choosing between these two platforms when establishing or expanding a high-throughput sequencing facility continues to be a difficult choice. For this reason, both instruments are described and contrasted in some detail. A third instrument to compete with the aforementioned two has been released by SpectruMedix (State College, PA). Because it has not yet been widely adopted, this instrument is discussed in less detail, but it may prove to add a third attractive option to the already difficult choice for a high-throughput sequencing platform. Lower-throughput CAE devices are also commercially available (8) but are not discussed in this article.

MEGABACE 1000 The MegaBACE 1000 is a 96-capillary array instrument based on four-color fluorescence detection through a scanning confocal imaging system. Capillaries have a 75-μm inner diameter and are 40 cm from injection to detection region. A 96-sample capacity is obtained with six replaceable arrays each consisting of 16 capillaries. Each capillary runs directly from the sample plate well, through a temperature-controlled chamber, into the confocal optical scanner where it joins the other 95 capillaries in forming a linear array. In this region, the capillary coating has been removed during manufacture to allow the confocal scanner optical access to the interior of the capillaries. Downstream of the scanning region, each set of 16 capillaries is bundled together and terminates in tubes of separation matrix or buffer. During operation, a confocal objective scans back and forth over the 96-capillary array, thus sampling fluorescence from each capillary by directing a laser beam of 488 nm/532 nm into the center of the capillary (through the capillary wall) and collecting any resulting fluorescence. The fluorescence is color separated through two dichroic mirrors and four filters and collected using photomultipliers. These filters and dichroic mirrors can be easily replaced to allow detection of fluorescence from different dye chemistries, including the recommended DYEnamic ET primer or dye terminator chemistries from Amersham Pharmacia Biotech, as well as Applied Biosystem's BigDye primer and terminator chemistries.

The separation matrix consists of linear polyacrylamide (LPA) provided by Amersham. Although LPA has excellent separation properties, it requires a special coating on the inside of the capillaries to reduce electroosmotic effects; it is possible that degradation of this coating is partially responsible for the limited lifetime of capillary bundles ($\sim$150–200 runs). The LPA is injected into the capillaries from the anode end under pressure from a compressed (1000 lb/in^2) nitrogen gas cylinder. Samples are loaded into the instrument in a single 96-well plate that is manually placed in a drawer at the bottom of the instrument. Once the drawer is closed, it is pneumatically lifted (with gas from a second cylinder at 120 lb/in^2) and brought against the ends of the 96 capillaries and 96 cathode electrodes. Loading of the matrix tubes is performed in similar fashion. The DNA samples are electrokinetically injected by applying 3–6 kV for 10–60 s (this is varied depending

on sample parameters). Because samples are injected directly from the sample plate into the capillary, each individual capillary and electrode must contact the surface of the sample. This requires each well of the sample plate to contain at least 5 μl of fluid.

ABI PRISM 3700 The ABI 3700 is also a capillary array instrument designed to process 96 samples per run, but it uses a single, replaceable array of 104 capillaries, of which 8 are spares to be used should some capillaries become permanently damaged or clogged. This usage of spare capillaries is permitted by the fact that the 3700 transfers samples from one of four sample plates to a loading bar before beginning electrokinetic injection. This both allows the geometry of the capillaries to remain as a single row of 104, rather than a two-dimensional array of 96, and allows the software to flag unusable capillaries and load samples intended for such capillaries in the spares instead. The 3700's use of a two-channel robotic pipettor also allows it to be loaded with up to four 96-well or 384-well plates to achieve a throughput of up to 8×96 samples in an unattended 24-h period. The capillary bundle in the 3700 is also held at constant temperature during the run and terminates at the detection region. The DNA bands in the 3700 are detected after they exit the capillaries in a sheath flow cuvette (13, 14). In this detection scheme, a constant flow of separation matrix is maintained around the outside of each capillary, thus hydrodynamically constraining the sample on exit from the end of the capillary long enough for it to be detected. Excitation of the dyes is performed by a fixed laser illuminating the entire row of capillary ends, whereas fluorescence detection is performed by chromatically separating the fluorescence emission by means of a concave spectrograph and recording the resulting two-dimensional image (wavelength in one dimension and capillary number in the other) with a cooled CCD camera. This detection system brings with it several advantages, including software-selectable "virtual" filter wavelengths to allow the use of any dye sets without having to replace physical elements in the instrument. It also provides for very high sensitivity and 100% duty cycle detection rather than collecting light from each capillary for only a small part of the duration of a scan, as in the MegaBACE. On the other hand, signal-intensity gradients, presumably due to attenuation of the excitation laser beam from one end of the array to the other, have been reported on this instrument (D Smailus, personal communication).

The 3700 uses POP-5 (Applied Biosystems, Foster City, CA) as a separation matrix. This is a polymer or combination of polymers that dynamically coats the capillary walls, thus allowing the use of bare capillaries and not suffering from signal-quality degradation from breakdown of the capillary coating. Despite this, reported capillary lifetimes are similar to the MegaBACE capillaries, although both can be extended through some "regeneration" treatments.

MEGABACE VS 3700 Reports from users (E Mardis, K McKernan, & D Smailus, personal communication) indicate that the MegaBACE produces about 100–200 more Phred20 bases per sample than the ABI 3700 but is more sensitive to failure from excess template or other contaminants. Usage of a template-suppression

reagent or cleaner templates can help reduce this problem. The 3700, on the other hand, has the major advantage of being much better automated and, with careful reagent usage, allowing up to eight runs unattended (K McKernan, personal communication). Unfortunately, this automation has contributed to some reported reliability problems, which may be taken care of by future software or hardware upgrades.

The choice between these two instruments continues to be difficult and somewhat dependent on the application. Table 1 lists some of the major sequencing centers and their inventory of sequencers. Most large academic centers have chosen the ABI 3700 as their main platform, primarily because of its more automated operation and higher tolerance to a variety of samples. It is widely acknowledged that the MegaBACE delivers longer reads, however, and many groups choose to use at least a few of these instruments to "finish" or fill in the gaps of the assembled sequence.

SPECTRUMEDIX SCE 9610 The latest entry to the high-throughput DNA sequencing market comes from SpectruMedix Corp (State College, PA). This company has recently released the SCE 9610, a fully automated, 96-capillary instrument designed to compete with the ABI 3700 and MegaBACE 1000. The 9610 can produce reads in excess of 600 bases in 2 h, with electrokinetic injection occurring directly from a 96-well sample plate and with dynamically coated capillaries. Any dyes emitting in the visible spectrum can be used on this instrument. Excitation is performed by side illumination of the capillary array, with the capillaries immersed in an index of refraction-matching fluid and contained between two reflective plates to guide the laser beam. Detection is by means of a CCD camera and a spectrograph to allow simultaneous detection of up to 30 colors. The instrument is capable of processing up to six plates unattended and provides other useful features, including full access to raw data files.

CAPILLARY TECHNOLOGY IN DEVELOPMENT Other capillary electrophoresis instruments have been developed but are not commercialized at this time. A

TABLE 1 Some of the major sequencing centers and their inventory of sequencers[a]

Center (date)	MD MegaBACE	ABI 3700
Wash. U. Genome Sequence Center (10/00)	13	93
Sanger Center (5/00)	35	104
Whitehead Institute (projected 12/00)	0	155
Baylor College of Medicine (10/00)	4	46
Joint Genome Institute (7/00)	84	7

[a]As of date shown in parentheses.

description of some of these follows. More detail on these, and on other instruments omitted here, can be found in other reviews (8, 18).

The RISA (Riken integrated sequence analyzer) developed at RIKEN (Tsukuba and Wako, Japan) is a 384-capillary instrument that uses a cross-linked matrix and disposable capillaries. The electrophoresis and detection instrument is one part of a larger system that includes two other instruments. The first of these, CAS, forms the 384-capillary bundles by cutting the capillaries to size, mounting them onto a base plate, and burning detection windows into the capillary coating. The second instrument, GVT, washes the capillaries and fills them with gel, at which point they are ready to be placed on the RISA instrument for sample loading and electrophoresis.

In an attempt to push the number of capillaries per instrument well beyond 96, Scherer et al (19) developed a 1000-capillary–capable instrument that, to date, has been tested with 128 capillaries and demonstrated reasonable (500 bp) read lengths. Continued development of this instrument is being performed at the Stanford DNA Sequence and Technology Center. Karger and coworkers, responsible for the introduction of LPA as a separation matrix (20), have demonstrated exceptional performance (21) with a single-capillary instrument, reaching 1000 bases in 55 min with accuracy between 98% and 99%. Dovichi and coworkers, originators of the sheath flow cuvette detection method used in the ABI 3700, have recently developed a very-high-sensitivity 5-capillary electrophoresis instrument (22, 23) for rapid separation and detection of small amounts of sample. This instrument is intended to fill a niche between single-capillary machines, such as the ABI 310, and the larger, 96-channel instruments. A 32-capillary version of this instrument has also been constructed.

Additional performance improvements in CAE may come from improved sample preparation and loading conditions. Work from the Karger group (24, 25) indicates that removing template from well desalted samples leads to improved performance. Usage of template-suppression reagents during loading can similarly improve reliability, particularly of instruments, such as the MegaBACE, that use LPA as a separation matrix. Xiong et al (26) have also shown that sample compression through artificially enhanced conductivity gradients in the injection region of the capillary can greatly increase read length of unpurified samples.

Microfabricated Capillary Arrays

Future performance improvements for CAE are likely to be achieved by miniaturization of CAE devices. This could reduce the cost of the devices, increase the number of capillaries available per instrument, and greatly reduce the separation time. Microfabricated devices are also more amenable to integration with the rest of the DNA sequencing process, leading to the possibility of sample preparation, thermal cycling, and electrophoretic separation all occurring on a single chip. [For more details on work toward integrated sequencing systems, see elsewhere (8, 27, 28).] This section concentrates on the latest developments in applications directed specifically at DNA sequencing.

Although theoretically similar to CAE, microfabricated devices have several possible advantages over conventional capillary instruments. One advantage is that a very large number of capillaries could be etched into a single wafer at low cost, making this technology amenable to highly parallel operation. Another significant advantage is that the manufacturing process used allows creation of geometries that would be nearly impossible with a standard capillary. An example of this is the T-channel injection geometry (Figure 1), where the main electrophoresis capillary is intersected, near the loading region, by an orthogonal capillary of the same dimensions. A modified form of electrokinetic injection is performed with this system by electrophoresing the sample into the short orthogonal capillary before turning the main capillary voltage on. The DNA migrates through the short capillary, filling the T intersection. When the main voltage is turned on, it is only the DNA in the T intersection that migrates down the main capillary. This allows the instrument designer to control the size of the injected DNA plug by controlling the dimensions of the intersection. Read lengths comparable to standard capillary sequencers have been obtained with this loading geometry and short (7–10 cm) capillaries.

Mathies and coworkers have fabricated capillary arrays on glass substrates using photolithography. The resulting etched glass channels are covered by a thermally bonded glass wafer with predrilled holes for access to the channels. This group has investigated various geometries and numbers of channels per chip (29–32). One of these (30) has been optimized for DNA sequencing. In this device, four capillaries of slightly different geometries were etched into a glass wafer of 4-in diameter. The 7-cm-long channels ranged from 20 μm to 50 μm in depth and from 70 μm to 130 μm in width, respectively. Electrokinetic injection of the sample is performed through a T-injector to enhance resolution. LPA was used as a separation matrix, and excitation and detection of the fluorescent dyes were performed through a four-color confocal detection system. Using this setup, 500 bases were resolved at over 99% accuracy in fewer than 20 min. A 96-channel device based on this principle would be capable of 150,000 bases/h. Shi et al (31) describe the possibility of building such a device based on previous work using a radial microplate design and high-speed rotary confocal scanner.

Ehrlich and coworkers have also used etched glass wafers prepared with methods similar to Mathies' to produce microfabricated sequencing devices. They have produced both short-capillary and longer-capillary devices. The short-capillary device consists of a single 11.5-cm-long capillary with a semicircular cross section of 40-μm radius. LPA was used as a separation matrix, and four-color fluorescence detection was used to identify the BigDye-labeled fragments. This device demonstrated an average read length of 505 bases at 99% accuracy in fewer than 30 min per run using real samples from human chromosome 17 (33). Optimization of this device (34) later led to read lengths up to 640 bases in 30 min at 98.5% accuracy using m13 mp18 standard. This group feels that this performance is limited by the properties of the separation matrix. To achieve longer reads, a similar device with capillaries 40 cm long, 40 μm deep, and 90 μm wide was constructed

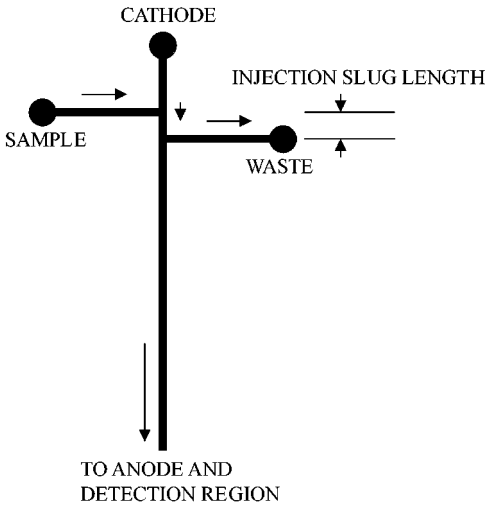

Figure 1 T-injector geometry for electrokinetic injection of sample in microfabricated capillary electrophoresis devices. The sample is electrophoresed from left to right and fills the T intersection. A part of the injected sample, of length equal to the channel overlap at the T intersection, is electrophoresed into the main capillary when current flow from the sample to the waste is terminated, and a potential is applied across the main cathode and anode.

(35) in etched glass plates measuring 25 cm $\times$ 50 cm. This device was used to demonstrate read lengths of 800 bases at 98% accuracy in 80 min.

A group at PE Biosystems, in collaboration with researchers from Alberta Microelectronic Corporation, has also been testing the feasibility of using large glass plates with multiple long, etched capillaries as a highly parallel sequencing platform (36). With procedures similar to those employed by the groups above, a device was constructed with 48 etched capillaries 50 cm in length and with semicircular cross sections 20–50 μm in length. In this case, POP-6 was used as a separation matrix, and reads of 640 bases with 98% accuracy in 2.5 h were reported. Although this system does not represent a performance increase over conventional CAE instruments, it may prove to be a path to less-expensive instrument manufacture and integration of sample processing with the electrophoresis instrument.

Molecular Dynamics has also recently reported (37) four-color DNA sequencing in 16-channel microfabricated CAE devices. These devices are also produced on glass wafers through photolithography and wet etching. Capillaries are 7–7.6 cm long, 110 μm wide, and 50 μm deep and are equipped with a twin T junction in the loading region. LPA is used for separation, and detection is by means of a four-color confocal scanner. Read lengths of up to 543 bases in 18 min with an accuracy of 99% are reported, with average reads ranging over 450 bases in 15 min.

It appears likely from these developments that the next generation of sequencing instruments will make use of some form of microfabricated, monolithic capillary array with substantially shorter run times. Aside from the expected benefits of chip-based CAE devices, including lower cost, higher throughput, and decreased run times, it appears (37) that these devices may also be less sensitive to high template concentrations in the samples and will likely be well suited to integration with small-volume thermal cycling technologies.

Free-Solution Electrophoresis

One major limitation of all current electrophoresis devices is that a compromise must be reached between longer reads and separation speed. In gel-based systems, it is unlikely that massive improvements in separation speed will occur without sacrificing read length. One possible approach to achieving such an improvement in separation time without sacrificing read length is free-solution electrophoresis. Noolandi (38, 39) and Mayer et al (40), among others (41, 42), have worked on the concept that although unlabelled DNA strands of different lengths migrate at the same velocity through an aqueous solution (not containing a separation matrix), appropriate labeling of the DNA strands with a friction-increasing molecule might allow strands of different lengths to be separated in a free solution under electric field–induced migration. Initial separations have been demonstrated (41) using a double-stranded 100-bp ladder (fragments from 100 bp to 1 kb) labeled with one or two streptavidin molecules. Detection of molecules was performed through ultraviolet (UV) absorbance. Separation resolution was sufficient to uniquely identify all 10 fragment lengths. More recent work (42) used the same method to resolve the first 100 bases of a DNA sequencing reaction by detection of single-stranded product fragments that were tagged with a streptavidin label. This separation was performed in only 18 min using 0.01% POP-6 polymer. Design of a larger, friction-generating molecule to allow longer reads is under way.

NONELECTROPHORETIC METHODS

In the 1980s and 1990s, several sequencing strategies were conceived that do not involve electrophoretic separation. Although many may never achieve the performance required to replace CAE for de novo DNA sequencing, these alternative sequencing strategies are developing into important diagnostic and research platforms.

Pyrosequencing

Pyrosequencing is a real-time sequencing strategy based on the release of pyrophosphate (PPi) during enzymatic DNA synthesis (43, 44). This process is initiated by annealing an oligonucleotide primer to a single-stranded DNA template in the presence of DNA polymerase. The first of four deoxynucleotide triphosphates

is then added to the reaction mixture. When the dNTP is the correct complement to the target strand, the DNA polymerase incorporates that nucleotide onto the 3′ end of the nascent strand, resulting in release of PPi at a concentration equimolar to nucleotide incorporation. Noncomplementary dNTPs result in little or no PPi release. The concentration of PPi in the system is monitored using a coupled enzymatic reaction in which ATP sulfurylase converts PPi to ATP, and the ATP subsequently drives conversion of luciferin to oxyluciferin by firefly luciferase with attendant emission of light (45, 46). The other three dNTPs are then each added independently in an iterative process. dATPαS is used in place of dATP because dATP is a substrate for luciferase, giving false-positive signals (43).

One requirement of this technology is that excess dNTP and ATP are removed before each successive dNTP addition. In some embodiments of the pyrosequencing technology (43), this step is accomplished by attaching the target sequence to a solid substrate that is washed between dNTP additions. This requires (*a*) replenishment of the enzyme mixture for each dNTP addition and (*b*) biotinylation of the target sequence to allow attachment of the single-stranded DNA template to streptavidin-coated paramagnetic beads. The solid-phase strategy has the advantage that inhibitory products do not accumulate during extension, and it is compatible with microfluidics and microarray technology (47). As an alternative, Ronaghi et al (44) developed a solution-phase assay in which they added apyrase to the polymerase reaction mixture to make a four-enzyme system. Apyrase catalyzes the degradation of ATP and dNTPs to yield nucleotide monophosphates and inorganic phosphate, resulting in a steady decline in light output to baseline. Once baseline is achieved, the next dNTP can be added and the cycle repeated numerous times without washing.

Initially, de novo sequencing was limited to about 30 nucleotides downstream of the annealed primer using the four-enzyme strategy. This limitation was due to asynchronous extension on some templates causing increased background and lower sequence-specific signal (44). This short read length was overcome to a degree when resequencing because dNTP addition could be programmed instead of cycling between four dNTPs. For example, Garcia et al (48) used pyrosequencing to detect mutations in exon 7 of the p53 gene in tumor samples from patients with nonmelanoma skin cancers. A single primer was annealed to exon 7 amplimers, and pyrosequencing was carried out by the programmed addition of nucleotides based on the wild-type consensus sequence of that gene. This permitted resequencing of all 109 bases of p53 exon 7 in approximately 100 min. More recently, addition of single-stranded DNA-binding protein was shown to give read lengths greater than 30 bases for de novo sequencing (49).

A commercial pyrosequencing instrument is available (PSQ^{TM}96, Pyrosequencing AB, Uppsala, Sweden) that uses a disposable ink-jet cartridge to deliver 200-nl aliquots of reagents into each of 96 wells. A lens array focuses luminescent light from the samples onto a CCD camera. This instrument has been used successfully for analysis of single nucleotide polymorphisms (SNPs) (50, 51); however, currently it is not widely used for de novo sequencing.

Sequencing by Hybridization

DNA arrays have become a fundamental tool in biology. In September 2000, over 30 commercial sources of arrays were available. Since then, major corporations, including Agilent, Motorola, Corning, Hitachi, and Mitsubishi Rayon, have announced plans to market DNA arrays, and in the year 2000 alone, hundreds of manuscripts were published in which DNA microarray development or application was a central focus. The principle underlying the technology is essentially the same as Southern blotting (52); however, scaling permits significant advantages in speed and cost (53, 54).

One early rationale for developing hybridization arrays was de novo sequencing (55, 56). As originally conceived, this strategy [sequencing by hybridization (SBH)] involved annealing a labeled unknown DNA fragment to a complete array of short oligonucleotides (e.g. all 65,336 combinations of 8-mers) and deciphering the unknown sequence from the annealing pattern. Over the past decade, SBH has largely been eclipsed by the use of DNA arrays for SNP and expression analysis. This is partly due to the amount of diagnostic or biological information to be gained per feature on the array. For example, expression monitoring of the entire human genome could be performed using a microarray composed of 100,000 gene-specific sequences (or possibly many fewer), whereas the same number of features would allow resequencing of only 25,000 bases (57). Another stumbling block for sequencing applications is the use of short oligonucleotide probes. These present such problems as ambiguous reads associated with repeat regions within the unknown target sequence, formation of secondary structures in some oligonucleotide probes that result in little or no detectable signal, and hybridization of oligonucleotides with single mismatches (false positives), which can be especially common at the terminal base pair. Each of these technical issues can be circumvented in expression arrays or in comparative sequencing by SBH (58, 59), but they cannot be avoided in de novo sequencing by SBH.

Massively Parallel Signature Sequencing

Massively parallel signature sequencing (MPSS) is an innovative sequencing strategy that was developed for expression analysis (60, 61). It combines the highly parallel nature of expression microarrays with the statistically robust digital output from gel-based sequencing of clones from representative cDNA libraries (62). Unlike most conventional microarrays, no prior knowledge of the target sequence is necessary for detection of a given cDNA.

The process begins by targeting members of an amplified cDNA library to the surface of microbeads using 32-mer tags composed of strings of eight 4-mer "words." The system is optimized so that each bead will carry approximately 100,000 cDNA copies of a unique mRNA (60). Microbeads carrying cDNAs are sorted by a fluorescence-activated cell sorter and then delivered to a micromachined flow cell, where they are immobilized against a vertical constriction to form a

planar, easily perfused array. Thus, each of about 1 million cDNA-coated beads can be analyzed in a single flow cell using conventional fluorescence microscopy.

To simultaneously read the sequence of the cDNA on each of the ~1 million beads, Brenner et al (60, 61) developed a protocol that couples a series of digestions by a restriction endonuclease (BbvI) with encoded adapters used to report successive cDNA-specific overhangs exposed by the enzyme. In the first step of this procedure, a short adapter containing the recognition sequence for BbvI is ligated to the end of each cDNA. BbvI (a type II restriction endonuclease) is unusual because its recognition sequence is eight nucleotides distal from the DNA segment that it cuts. Thus, when BbvI enzyme is supplied to the bead array, it binds to all the cDNAs at the common recognition sequence, but it cuts at four-nucleotide segments downstream that are cDNA-specific. These overhangs are then annealed and ligated to a panel of adapters, each of which incorporates three features: (*a*) another copy of the BbvI recognition sequence; (*b*) a four-nucleotide overhang that can be annealed to overhangs on the cDNAs; and (*c*) a 10-nucleotide overhang that serves to encode the adapter. In the next step the encoded adapters on each bead are identified using 16 fluorescently tagged "decoder oligonucleotides" that are annealed to the encoded adapters and read in succession. Only 16 decoder oligonucleotides are needed to identify 1024 possible encoded adapters, as each four-nucleotide overhang is read one base at a time. After all 16 decoder molecules have been tested, the cDNA on each bead will have bound to four decoder molecules, thus identifying the sequence of the four-nucleotide overhang on that bead. The sequences are then cut again by BbvI, exposing the next four nucleotides of the cDNA, which are annealed to encoded adapters and decoder oligonucleotides as before. This cycle is repeated up to five times, yielding 20 bases of sequence per cDNA. Brenner et al (61) do not state whether readable sequence per strand can be extended beyond 20 bases. It is possible that asynchrony is an issue for longer read lengths, as is true for pyrosequencing (44).

The rate of de novo sequence acquisition by MPSS is impressive. A typical analysis of 20 bases per bead using 1 million beads can be completed in a few days using a single apparatus. Given the short read length, this approach is best suited to expression analysis. Application of MPSS to de novo genomic sequencing would require a substantial increase in read length per strand.

Sequencing by MALDI-TOF Mass Spectrometry

Several technologies have emerged over the past decade that facilitate mass spectrometry of macromolecules. These include electrospray ionization mass spectrometry, Fourier transform ion cyclotron resonance mass spectrometry, and matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS). This section focuses on MALDI-TOF-MS, as it has been involved in most DNA sequence determination by mass spectrometry.

MALDI was first coupled to time-of-flight mass spectrometry in 1988 by Karas & Hillenkamp (63). Their remarkable innovation was that virtually any macromolecule could be desorbed as an intact gas-phase ion by embedding it in

the crystal of a low-molecular-weight molecule that strongly absorbs energy from a pulse of laser light. Prior to this landmark paper, laser desorption mass spectrometry was limited to peptides with specific volatility or photoabsorption properties (64).

Although originally applied to analysis of protein samples, MALDI-TOF-MS is now widely used for oligonucleotides and DNA as well. The essential features of MALDI-TOF-MS DNA analysis are summarized as follows. The DNA sample is typically dried at room temperature on a flat surface in a matrix of 3-hydroxypicolic acid. The 3-hydroxypicolic acid matrix serves the critical purpose of absorbing UV light while interacting very little with DNA (65, 66). The sample is then treated with a short pulse of UV laser light that is absorbed by the 3-hydroxypicolic acid, causing ablation of DNA ions into the gas phase. The DNA ions are generally monovalent and intact (hence the term soft ionization). After a specified time delay (67), the charged gas-phase DNA molecules are extracted by a high-voltage pulse and accelerated in an electric field so that they attain a common kinetic energy. They are subsequently passed into a flight tube approximately 1 m long. Under vacuum at a common kinetic energy, the relative time required for a given molecule to travel the flight path is dependent on its mass. At the end of the flight tube, the molecules collide with an ion-to-electron conversion detector, thus registering the TOF from the original laser pulse (Δt_m). The mass of a given analyte can then be calculated from the relationship $m = 2qV\Delta t_m^2/2l^2$, where m is the mass of the ion, q is the charge on the ion, V is the voltage used to accelerate the ion, and l is the length of the flight tube. In practice, internal standards are often relied on to confirm peak identities.

DNA sequencing by MALDI-TOF-MS was first explored in the early 1990s using DNA ladders generated by the Sanger dideoxy-terminator method (for a review, see 68). In essence, MALDI-TOF-MS DNA sequencing adopted the enzymology and nucleic acid chemistry established for conventional sequencing; however, strand separation through gels was replaced by mass-dependent strand separation of gas-phase ions (69). This strategy is attractive because of three inherent features of MALDI-TOF-MS: (*a*) Separation by mass along the flight path permits simultaneous analysis of many DNA strands in a heterogeneous mixture; (*b*) MALDI-TOF-MS DNA analysis is fast (one spectrum can be acquired in about 1 ms, and integration of spectra from 100 or so repeated laser shots can be acquired in a matter of seconds); and (*c*) charge-to-mass ratio measured by MALDI-TOF-MS is an inherent property of a given DNA strand and is, therefore, not affected by secondary structure, as is true for gel-based methods and SBH (69).

The state of the art in MALDI-TOF-MS sequencing was summarized in 1996 by Smith (70). In his commentary, he pointed out that there was "no mad stampede" of large-scale sequencing centers to replace their electrophoretic systems with MALDI-TOF-MS (this assessment preceded the advent of high-throughput capillary sequencing devices). At that time, the maximum read length for a DNA ladder in a single spot measured by MALDI-TOF-MS was 89 nucleotides (nt) (65), which was a significant disadvantage for de novo sequencing. It was postulated

that this MALDI-TOF-MS read length was limited by strand fragmentation (70) and salt-adduct formation (71). In the former case, the underlying mechanism was believed to be strand scission following depurination caused by protonation of the nucleobase (72, 73). For example, oligomers composed of deoxythymidylic acid yielded readable spectra up to 100 nt in length, whereas oligomers composed of a mixture of G, A, T, and C nucleotides gave readable spectra to only 26 nt in length (74). Further, in 1996, MALDI-TOF-MS sequencing had not fulfilled its promise of being high speed. This was in part because samples were delivered to the probe surface in relatively large volumes (0.3 μl) using hand pipets so that the dried spot was significantly larger than the $\sim$0.1-mm laser beam used to desorb the analyte. This meant that sensitivity was compromised and that the operator had to manually hunt through the spot to generate a high-quality spectrum—a clear disadvantage for a technology aiming for speed and automation.

Since 1996, the speed of MALDI-TOF-MS DNA sequence analysis has advanced markedly, in part because of the reproducible delivery of samples to spots on flat arrays. For example, automated piezoelectric delivery of nanoliter sample volumes to the surface of 1-in^2 silicon chips resulted in 100 uniform spots of a 25-nt oligomer (75). Because each spot was about equal in diameter to the laser beam used to vaporize the sample ($\sim$0.1 mm), similar spectra were achieved with as little as 8 fmol of oligonucleotide. This meant there was no need to manually scan a given spot to acquire a robust signal and that automated analysis from spot to spot was practical. This is illustrated by a MALDI-TOF-MS DNA array developed by Sequenom Corporation in San Diego, CA. In their MassArray technology, short sequence-specific DNA ladders are synthesized in parallel and then simultaneously loaded in nanoliter volumes onto matrix spots arranged in an array on a silicon chip. Each chip may contain up to 384 such elements, and 10 of these chips (3840 samples) may be loaded onto a MALDI-TOF-MS for an unattended run. Becker and colleagues at Gene Trace Systems have also developed a MALDI-TOF mass spectrometer that automatically locates sample spots and searches a particular mass range without operator intervention (76). This permits analysis of a given sample in a few seconds, and thousands of DNA samples per day.

In contrast to these advances in speed and automation, the sequence read length that can be achieved by MALDI-TOF-MS has changed little since 1996. Analysis of DNA ladders generated by the Sanger method has been extended to 100 bases in two laboratories (77, 78), but this read length is not routine. Tang et al in Wisconsin have addressed the problem of nucleobase-dependent strand fragmentation by replacing the 2′-hydrogen on the deoxyribose sugar with electron withdrawing groups (e.g. 2′-fluorodeoxycytidine nucleoside) (79). As anticipated, oligonucleotides incorporating the cytidine analogs partially or completely prevented fragmentation at the substituted position. MALDI-TOF analysis of mixtures of these substituted oligonucleotides gave extended mass range and mass resolution relative to unsubstituted oligonucleotide mixtures. The authors postulate that chain-termination sequencing reactions using 2′-fluoro-dNTPs could give longer

reads by MALDI-TOF-MS, but this will require polymerases that can efficiently incorporate the modified substrate.

In summary, advances in automation have made MALDI-TOF-MS a competitive tool for high-speed DNA sequence analysis. The technology is especially strong at precise resequencing using short DNA fragments. This is reflected in the recent emphasis on SNP analysis by MALD-TOF-MS (80–86). In contrast, there is currently no prospect of reaching 600 nucleotides per reaction, as is common with high-throughput capillary devices. Therefore, MALDI-TOF-MS is unlikely to be widely used in de novo genomic sequencing the foreseeable future.

Single-Molecule Methods

Since the advent of the Sanger dideoxy sequencing method in 1977, advances in instrumentation have occurred that permit detection and functional analysis of single molecules. Among these advances are atomic-force microscopy (87), single-molecule fluorescence spectroscopy (88, 89), and single-channel-current measurements (90–92). Each of these technologies has been used to analyze physical properties of individual DNA strands, leading to suggestions that their use could be extended to single-nucleotide resolution and rapid DNA sequence analysis (93–97). Although none of these approaches has enjoyed even minimal success at DNA sequencing, they merit discussion because if successfully implemented, they could yield large improvements in speed, read length, and sensitivity.

ION CHANNELS In the 1970s, ionic currents through single-membrane protein channels were observed using polypeptides intercalated in artificial planar membranes (90, 91). The same basic principles drew wide interest when it was demonstrated that patches of biological membranes sealed at the ends of micrometer-diameter glass tips could be used to study individual ion channels from higher organisms (92).

It was recently proposed that such a channel could be used to analyze the base sequence of one DNA strand during translocation under an applied field (93, 98). The advantages to this approach are impressive: single-molecule sensitivity that could eliminate laborious cloning and replication steps; read lengths limited only by the size of DNA that can be managed without shearing; and high-speed analysis approaching 1000 bases per s. The daunting technical requirements of such a device are equally impressive. These requirements fall into two categories: (*a*) a pore that ensures single-file, unidirectional transport of DNA strands across a defined aperture at nanometer precision; and (*b*) a detector that can register very subtle structural or chemical differences between adjacent unmodified nucleotides separated by only a few angstroms.

To date, the available candidates for such a nanoscale pore are biological channels. Most of these form pores that are much narrower than the average diameter of B-form dsDNA or ssDNA and, therefore, could not fulfill the first requirement. One exception is the pore formed by α-hemolysin, a 33-kDa protein secreted by *Staphylococcus aureus*. Figure 2 shows a cross section of the α-hemolysin channel

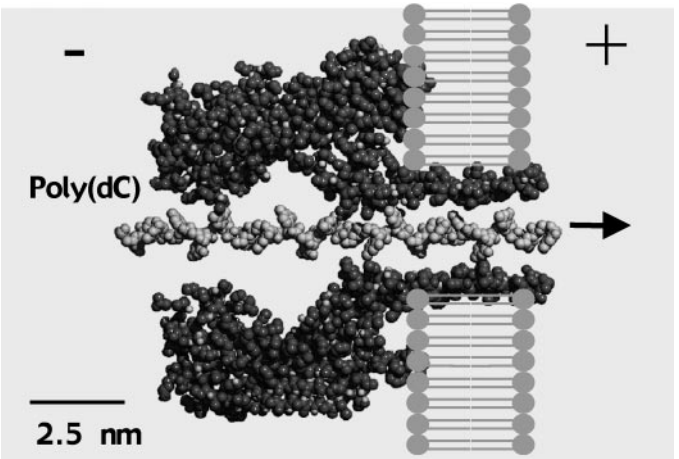

Figure 2 Cross section of the α-hemolysin channel embedded in a lipid bilayer. A single strand of poly(dC) is shown inside the pore as it is translocated under an applied voltage. The image of α-hemolysin was redrawn based on coordinates provided by Song et al (99).

at 1.9-Å resolution as revealed by X-ray crystallography (99). This heptamer assembly is about 100 Å in total length, with a relatively narrow beta-barrel segment believed to span the membrane barrier, and a broader segment (the "mushroom cap") that extends into one of the aqueous compartments. The mouth of the channel is about 26 Å in diameter—wide enough to accommodate B-form duplex DNA. The pore then widens into a vestibule that abruptly narrows to a limiting aperture of 15 Å at about the membrane-solution interface. This limiting aperture is just larger in diameter than an extended single strand of DNA. At pH 8.0 in 1 M KCl, current through the α-hemolysin pore is robust (120 pA at a 120-mV applied potential and room temperature) and spontaneous gating is rare (root mean square noise <1 pA at a 5-kHz bandwidth).

In a landmark paper, Kasianowicz et al (93) used reduction of this otherwise unimpeded ionic current to detect individual single-stranded DNA and RNA molecules as they traversed the α-hemolysin pore under an applied field. A typical event caused by translocation of a poly(U) 150-mer is shown in Figure 3. As predicted by the crystallographic model, ssDNA (but not dsDNA) caused 85% blockades characteristic of translocation events and could be amplified by PCR from the *trans* compartment. The polynucleotide blockade durations were strand-length dependent, indicating that DNA traverses the pore as an extended strand. Thus, the requirement of single-file transport across a nanometer-scale aperture is fulfilled by this prototype channel. It remains unproven, however, whether DNA translocation through α-hemolysin is unidirectional, or whether thermal fluctuations could cause the strand to randomly diffuse backward against the applied electric field.

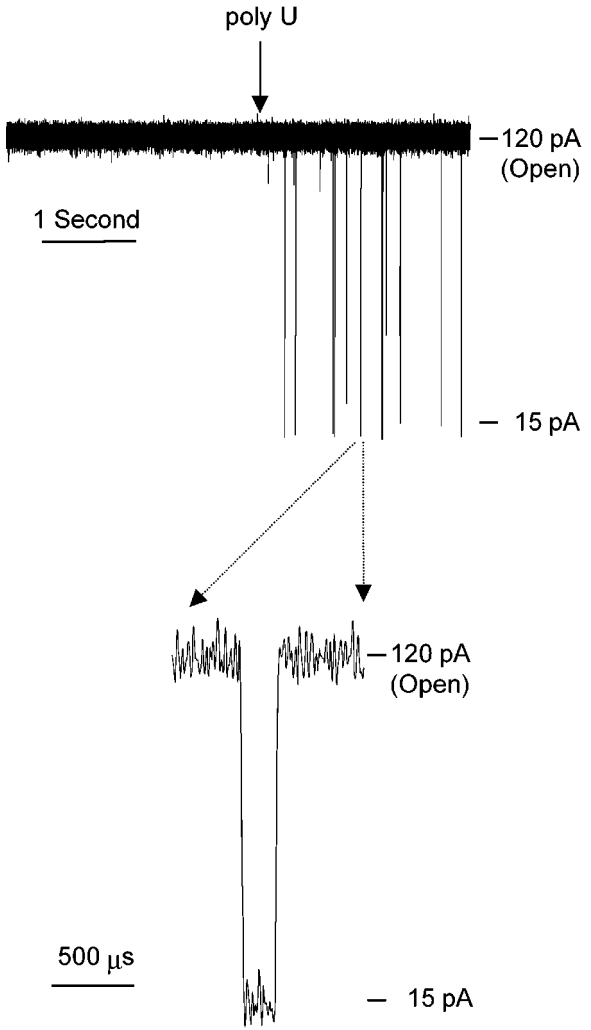

Figure 3 Blockade of ionic current through the α-hemolysin pore during translocation of poly(U) RNA homopolymers. The poly(U) strands averaged 150 nt in length. The applied potential was 120 mV at room temperature. The buffer was 1 M KCl and 10 mM HEPES/KOH at pH 8.0.

Modulation of ionic current is an adequate detector that can report the base composition of polynucleotides at low resolution. For example, individual RNA homopolymers (100) and individual DNA block copolymers of different base compositions (101) can be distinguished from one another in two-dimensional plots of blockade duration vs blockade amplitude. These diagnostic signatures are due to nucleotide-dependent secondary structures adopted by the polymers rather

than to base-specific impedance of the pore. A 30-residue abasic insert in the middle of a synthetic poly(dC) oligomer also causes a reproducible current transition when threaded through the α-hemolysin pore (102). This current transition is probably caused by a specific difference between the abasic monomer and deoxycytidylic acid rather than by secondary structure.

Detection of monovalent ion current through the α-hemolysin pore is not likely to yield DNA sequence at single-nucleotide resolution, for two reasons. First, the narrowest portion of the channel pore is 50 Å long, meaning that approximately seven nucleotides occupy that space at a given instant. Each of those seven nucleotides would contribute to resistance against ionic current, thus obscuring the influence of any single nucleotide. Second, the number of monovalent ions flowing through the narrowest segment of the pore during ssDNA translocation is quite small. For example, the largest current difference published to date is that between long homopolymers of poly(C) and poly(A) RNA (100). When this current difference is calculated per nucleotide within the strand, it is at most 100 monovalent ions—a level that cannot be discerned (103).

Improvements in resolution hinge in part on slowing DNA translocation. A promising approach would use duplex DNA as a brake from which one single-stranded complement could be systematically threaded into the pore (98). This could be achieved either enzymatically or by using the electric field across the pore to unzip duplex DNA. For example, Vercoutere et al (104) have shown that synthetic DNA molecules with internal duplex hairpins are translocated through the α-hemolysin pore at a 120-mV applied potential and room temperature. For these hairpins, the rate of translocation is slowed 5- to 10-fold per base pair relative to random sequence, presumably due to the nonequilibrium, sequential dissociation of each base pair as it is pulled toward the pore aperture. Based on earlier work with DNA block copolymers (101), it is reasonable to expect that this difference would be accentuated at cooler temperatures.

Improvements in resolution also hinge on more sensitive detectors coupled to fabricated nanopores. One such device is being developed at Harvard University. In that scheme (98), ion beams are used to etch individual nanometer-scale holes in solid state membranes. Other engineered nanopores are under development, including etched mica (105), modified protein channels (106), and pores formed by transmembrane assemblies of *p*-octiphenyls with leucine/lysine side chains (107). Currently, none of these alternatives has been shown to efficiently translocate DNA.

ATOMIC-FORCE MICROSCOPY Atomic-force microscopy (AFM) was first described in 1986 (87). The core of this technology is a nanoscopic tip attached to the end of a flexible cantilever. The distance between the AFM tip and a target surface is controlled by a piezoelectric device. When the tip approaches and then is scanned across the target surface, interactions between the tip and the surface cause the cantilever to deflect. The measured deflection (detected by a laser reflected from the surface of the cantilever) is used to reposition the tip to avoid contact with the surface and resulting tip damage. This process generates a topographic map of the

surface. Resolution by AFM in this mode depends in part on the radius and shape of the tip (108, 109). Typical commercial Si or Si_3N_4 tips may have tip radii as small as 10 nm (110). In an early study (111), AFM was used to scan plasmid, and Hae III–digested ϕX174 DNA adsorbed to a cationic lipid bilayer surface in aqueous buffer at room temperature. When the DNA was closely packed, high-resolution images revealed a right-handed double helix with the expected pitch of 3.4 nm. On this basis, the authors speculated that single-nucleotide resolution of oriented, single-stranded DNA or RNA might be achieved using more refined tips.

Although this ambitious goal remains elusive, measurable improvements in AFM resolution have been achieved. For example, Wong et al (110) prepared single-walled carbon nanotubes (112) and attached these to Au-coated Si cantilevers. Using monodisperse Au nanocrystals as standard targets, they found that the tips of their single-walled carbon nanotubes had radii 3–6 nm in diameter and that the lateral resolution of these tips in tapping mode was 70% better than the resolution of conventional Si tips. AFM using single-walled carbon nanotube tips 3 nm in radius was later successfully applied to haplotyping using gene-specific oligonucleotide 10-mers labeled with biotin-streptavidin or with a fluorophore (IRD800) (94). In this case, discrimination relied on the size and shape of the two labels rather than on the structure of DNA per se, consistent with an earlier AFM study that detected a modified streptavidin bound to the end of biotinylated DNA (113). Speed and resolution of AFM for scanning DNA samples is likely to improve further because of the recent fabrication of nanotubes with radii as small as 0.25 nm (114) and because of the development of multiple tip arrays (115–117).

AFM in a force-measuring mode (118) is one of several strategies used to measure the mechanical properties of DNA (119–124). This involves attaching one end of the polymer to an immobile surface (e.g. by drying onto gold) and the other end to the AFM tip. As the tip and the surface are separated using a piezoelectric device, deflection of the AFM cantilever generates a force-extension curve characteristic of the DNA molecule. The stretched DNA may be single stranded, double stranded with both ends of each molecule attached to the tip and the surface, or double stranded with only one end attached to the AFM tip, permitting "unzipping" of the duplex. The latter arrangement was employed by Rief et al (125) to measure the unbinding forces for G-C and A-T pairs. In the case of G-C pairs, this force was found to be 20 pN, and in the case of A-T pairs, it was found to be 10 pN. This is in close agreement with a similar study in which the base-pair unbinding force was estimated based on average G-C content of regions of bacteriophage λ DNA that were unzipped using a fine glass needle (95).

It has been proposed that this unbinding force could be useful in large-scale sequencing (95, 125, 126). However, currently the best resolution achieved by mechanical unzipping of DNA is approximately 500 bases, using the glass-needle device operating at 20–40 bp/s (95).

SINGLE-MOLECULE FLUORESCENCE MICROSCOPY Several laboratories have proposed to sequence individual DNA molecules by fluorescence microscopy (97, 127, 128). In a typical protocol (97), a DNA polymerase is used to copy a target template

using fluorescently tagged nucleotides as substrate. Only two of the nucleotides need to be fluorescent per reaction, provided the labels can be permuted among the four possibilities and examined separately. The fluorescently tagged duplex DNA is anchored in an aqueous stream, and the labeled strand is digested from the 3′ end by an exonuclease (e.g. exonuclease III). The rate of digestion and the flow rate are regulated so that the fluorescent nucleotides cut from the labeled strand are delivered sequentially to an approximately 1-pl volume. In this small volume, background fluorescence is minimized (129), permitting detection of bursts of 10–100 photons from the fluorescent nucleotides as they are excited by a laser of appropriate wavelength focused on the probe volume. It is postulated that this approach could read hundreds of bases per second with read lengths of 10–30 kb (129, 130).

Some of the requirements of this strategy have been met since it was conceived, i.e. modified nucleotides can be incorporated into DNA up to 2500 bases in length (see 130), and single fluorophores can be detected and identified at reasonable efficiency (131–137). However, the sequence of a single strand has yet to be read, and the practical rate of exonuclease-dependent fluorescent nucleotide detection is significantly slower than projected. These points are illustrated in a recent study by Werner et al (130). A 45-nt dsDNA was prepared with biotin at one 5′ end, and with tetramethylrhodamine dUMP and rhodamine 6G-dCMP each incorporated at one position along the reverse strand. Approximately 4500 copies of this DNA construct were attached to a single microsphere held in a 0.5-cm/s flow by a focused laser beam. Digestion by exonuclease III at 25°C delivered 1 nt per strand per s to the excitation laser probe volume. Based on fluorescence burst size and fluorescence lifetime, an individual fluorophore traversing the probe volume could be identified with 90% confidence. To achieve 99.9% certainty for a base call, the authors estimate that a given sequence must be reread 12 times. This, coupled with the six separate labeling reactions that are needed to examine each permutation of two labeled nucleotides, gives a total of 72 individual strands that would have to be analyzed to yield reliable sequence data. Using a processive enzyme such as exonuclease I, fluorescent nucleotides could be cut and delivered to the excitation laser at approximately 50/s for an overall sequencing rate of 1 nt/s. This rate is approximately equal to a rate already achieved by the Mathies group using fabricated capillary arrays on glass substrates (30). Thus, for the single-molecule fluorescence microscopy strategy to become competitive, either substantial improvements in base-call precision or read length would have to be demonstrated.

CONCLUSIONS

Electrophoretic methods yield by far the longest reads of any current sequencing strategies, and their performance continues to improve. They are unlikely to be supplanted for large-scale de novo sequencing in the foreseeable future. Microfabricated devices coupled with integrated sample processing may constitute the next

generation of sequencing instruments. This type of development could increase throughput at the largest sequencing centers by 10- to 100-fold. Highly parallel strategies including hybridization microarrays, MPSS, MALDI-TOF-MS, and pyrosequencing are well suited to SNP and expression analysis. However, none of these platforms is likely to contribute to large-scale de novo genomic sequencing because of their short read lengths. Sequence analysis and assembly is already a major bottleneck to genome sequencing (138), and any technology that yields short reads is at a serious disadvantage for this application. Single-molecule methods, although still in their infancy, are potential candidates for a major revolution in the way de novo sequencing is performed.

Visit the Annual Reviews home page at www.AnnualReviews.org

LITERATURE CITED

1. Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463–67
2. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, et al. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* 321:674–79
3. Smith LM, Fung S, Hunkapiller MW, Hunkapiller TJ, Hood L. 1985. The synthesis of oligonucleotides containing an aliphatic amino group at the 5′ terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Res.* 13:2399–412
4. Watson JD. 1990. The human genome project: past, present, and future. *Science* 248(4951):44–49
5. Cantor CR. 1990. Orchestrating the Human Genome Project. *Science* 248(4951): 49–51
6. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8(3):175–85
7. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8(3): 186–94
8. Meldrum D. 2000. Automation for genomics. Part Two: Sequencers, microarrays, and future trends. *Genome Res.* 10: 1288–303
9. Huang GM. 1999. High-throughput DNA sequencing: a genomic data manufacturing process. *DNA Seq.* 10:149–53
10. Kostichka AJ, Marchbanks ML, Brumley RL Jr, Drossman H, Smith LM. 1992. High speed automated DNA sequencing in ultrathin slab gels. *Biotechnology* 10(1):78–81
11. Erfle H, Ventzki R, Voss H, Rechmann S, Benes V, et al. 1997. Simultaneous loading of 200 sample lanes for DNA sequencing on vertical and horizontal, standard and ultrathin gels. *Nucleic Acids Res.* 25:2229–30
12. Huang XC, Quesada MA, Mathies RA. 1992. DNA sequencing using capillary array electrophoresis. *Anal. Chem.* 64(18):2149–54
13. Swerdlow H, Zhang JZ, Chen DY, Harke HR, Grey R, et al. 1991. Three DNA sequencing methods using capillary gel electrophoresis and laser-induced fluorescence. *Anal. Chem.* 63(24):2835–41
14. Swerdlow H, Jones BJ, Wittwer CT. 1997. Fully automated DNA reaction and analysis in a fluidic capillary instrument. *Anal. Chem.* 69(5):848–55
15. Deleted in proof

16. Deleted in proof
17. Deleted in proof
18. Carrilho E. 2000. DNA sequencing by capillary array electrophoresis and microfabricated array systems. *Electrophoresis* 21(1):55–65
19. Scherer JR, Kheterpal I, Radhakrishnan A, Ja WW, Mathies RA. 1999. Ultrahigh throughput rotary capillary array electrophoresis scanner for fluorescent DNA sequencing and analysis. *Electrophoresis* 20(7):1508–17
20. Ruiz-Martinez MC, Berka J, Belenkii A, Foret F, Miller AW, Karger BL. 1993. DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection. *Anal. Chem.* 65(20):2851–58
21. Salas-Solano O, Carrilho E, Kotler L, Miller AW, Goetzinger W, et al. 1998. Routine DNA sequencing of 1000 bases in less than one hour by capillary electrophoresis with replaceable linear polyacrylamide solutions. *Anal. Chem.* 70(19): 3996–4003
22. Zhang J, Voss KO, Shaw DF, Roos KP, Lewis DF, et al. 1999. A multiple-capillary electrophoresis system for small-scale DNA sequencing and analysis. *Nucleic Acids Res.* 27(24):e36
23. Crabtree HJ, Bay SJ, Lewis DF, Zhang J, Coulson LD, et al. 2000. Construction and evaluation of a capillary array DNA sequencer based on a micromachined sheath-flow cuvette. *Electrophoresis* 21(7):1329–35
24. Ruiz-Martinez MC, Salas-Solano O, Carrilho E, Kotler L, Karger BL. 1998. A sample purification method for rugged and high-performance DNA sequencing by capillary electrophoresis using replaceable polymer solutions. A. Development of the cleanup protocol. *Anal. Chem.* 70(8):1516–27
25. Salas-Solano O, Ruiz-Martinez MC, Carrilho E, Kotler L, Karger BL. 1998. A sample purification method for rugged and high-performance DNA sequencing by capillary electrophoresis using replaceable polymer solutions. B. Quantitative determination of the role of sample matrix components on sequencing analysis. *Anal. Chem.* 70(8):1528–35
26. Xiong Y, Park SR, Swerdlow H. 1998. Base stacking: pH-mediated on-column sample concentration for capillary DNA sequencing. *Anal. Chem.* 70(17):3605–11
27. Schmalzing D, Koutny L, Salas-Solano O, Adourian A, Matsudaira P, Ehrlich D. 1999. Recent developments in DNA sequencing by capillary and microdevice electrophoresis. *Electrophoresis* 20(15–16):3066–77
28. Dolnik V. 1999. DNA sequencing by capillary electrophoresis. *J. Biochem. Biophys. Methods* 41(2–3):103–19
29. Woolley AT, Lao K, Glazer AN, Mathies RA. 1998. Capillary electrophoresis chips with integrated electrochemical detection. *Anal. Chem.* 70(4):684–88
30. Liu S, Shi Y, Ja WW, Mathies RA. 1999. Optimization of high-speed DNA sequencing on microfabricated capillary electrophoresis channels. *Anal. Chem.* 71(3):566–73
31. Shi Y, Simpson PC, Scherer JR, Wexler D, Skibola C, et al. 1999. Radial capillary array electrophoresis microplate and scanner for high-performance nucleic acid analysis. *Anal. Chem.* 71(23):5354–61
32. Paegel BM, Hutt LD, Simpson PC, Mathies RA. 2000. Turn geometry for minimizing band broadening in microfabricated capillary electrophoresis channels. *Anal. Chem.* 72(14):3030–37
33. Schmalzing D, Tsao N, Koutny L, Chisholm D, Srivastava A, et al. 1999. Toward real-world sequencing by microdevice. *Genome Res.* 9(9):853–58
34. Salas-Solano O, Schmalzing D, Koutny L, Buonocore S, Adourian A, et al. 2000. Optimization of high-performance

DNA sequencing on short microfabricated electrophoretic devices. *Anal. Chem.* 72(14):3129–37
35. Koutny L, Schmalzing D, Salas-Solano O, El-Difrawy S, Adourian A, et al. 2000. Eight hundred-base sequencing in a microfabricated electrophoretic device. *Anal. Chem.* 72(14):3388–91
36. Backhouse C, Caamano M, Oaks F, Nordman E, Carrillo A, et al. 2000. DNA sequencing in a monolithic microchannel device. *Electrophoresis* 21(1):150–56
37. Liu S, Ren H, Gao Q, Roach DJ, Loder RT Jr, et al. 2000. Automated parallel DNA sequencing on multiple channel microchips. *Proc. Natl. Acad. Sci. USA* 97(10):5369–74
38. Noolandi J. 1992. A new concept for sequencing DNA by capillary electrophoresis. *Electrophoresis* 13(6):394–95
39. Noolandi J. 1993. A new concept for separating nucleic acids by electrophoresis in solution using hybrid synthetic end labelled-nucleic acid molecules. *Electrophoresis* 14(8):680–81
40. Mayer P, Slater GW, Drouin G. 1994. Theory of DNA sequencing using free-solution electrophoresis of protein-DNA complexes. *Anal. Chem.* 66:1777–80
41. Heller C, Slater GW, Mayer P, Dovichi N, Pinto D, et al. 1998. Free-solution electrophoresis of DNA. *J. Chromatogr. A* 806:113–21
42. Ren H, Karger AE, Oaks F, Menchen S, Slater GW, Drouin G. 1999. Separating DNA sequencing fragments without a sieving matrix. *Electrophoresis* 20(12):2501–9
43. Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 242:84–89
44. Ronaghi M, Uhlen M, Nyren P. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281:363–65
45. Nyren P. 1987. Enzymatic method for continuous monitoring of DNA polymerase activity. *Anal. Biochem.* 167:235–38
46. Nyren P, Pettersson B, Uhlen M. 1993. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal. Biochem.* 208:171–75
47. Ronaghi M. 2001. Pyrosequencing sheds light on DNA sequencing. *Genome Res.* 11:3–11
48. Garcia CA, Ahmadian A, Gharizadeh B, Lundeberg J, Ronaghi M, Nyren P. 2000. Mutation detection by pyrosequencing: sequencing of exons 5–8 of the p53 tumor suppressor gene. *Gene* 253:249–57
49. Ronaghi M. 2000. Improved performance of pyrosequencing using single-stranded DNA-binding protein. *Anal. Biochem.* 286:282–88
50. Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P, et al. 2000. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* 280:103–10
51. Alderborn A, Kristofferson A, Hammerling U. 2000. Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Res.* 10:1249–58
52. Southern EM. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–17
53. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* 93:10614–19
54. Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–70
55. Drmanac R, Drmanac S, Labat I, Crkvenjakov R, Vicentic A, Gemmell A. 1992. Sequencing by hybridization: towards an automated sequencing of one million

M13 clones arrayed on membranes. *Electrophoresis* 13:566–73
56. Drmanac R, Drmanac S, Strezoska Z, Paunesku T, Labat I, et al. 1993. DNA sequence determination by hybridization: a strategy for efficient large-scale sequencing. *Science* 260:1649–52; Erratum. 1994. *Science* 163(5147):596
57. Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW. 1998. Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* 16:301–6
58. Drmanac S, Kita D, Labat I, Hauser B, Schmidt C, et al. 1998. Accurate sequencing by hybridization for DNA diagnostics and individual genomics. *Nat. Biotechnol.* 16:54–58
59. Chee M, Yang R, Hubbell E, Berno A, Huang XC, et al. 1996. Accessing genetic information with high-density DNA arrays. *Science* 274:610–14
60. Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, et al. 2000. In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc. Natl. Acad. Sci. USA* 97:1665–70
61. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, et al. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* 18:630–34
62. Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, et al. 1992. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat. Genet.* 2:173–79
63. Karas M, Hillenkamp F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* 60:2299–301
64. Beavis RC, Chait BT. 1996. Matrix-assisted laser desorption ionization mass-spectrometry of proteins. *Methods Enzymol.* 270:519–51
65. Wu KJ, Shaler TA, Becker CH. 1994. Time-of-flight mass spectrometry of underivatized single-stranded DNA oligomers by matrix-assisted laser desorption. *Anal. Chem.* 66:1637–45
66. Wu KJ, Steding A, Becker CH. 1993. Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix. *Rapid Commun. Mass Spectrom.* 7:142–46
67. Roskey MT, Juhasz P, Smirnov IP, Takach EJ, Martin SA, Haff LA. 1996. DNA sequencing by delayed extraction matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Proc. Natl. Acad. Sci. USA* 93:4724–29
68. Guo BC. 1999. Mass spectrometry in DNA analysis. *Anal. Chem.* 71:333–37R
69. Cantor CR, Tang K, Graber JH, Maloney M, Fu DJ, et al. 1997. DNA sequencing after the Human Genome Project. *Nucleosides Nucleotides* 16:591–98
70. Smith LM. 1996. Sequence from spectrometry: a realistic prospect? *Nat. Biotechnol.* 14:1084–87
71. Lin H, Hunter JM, Becker CH. 1999. Laser desorption of DNA oligomers larger than one kilobase from cooled 4-nitrophenol. *Rapid Commun. Mass Spectrom.* 13:2335–40
72. Nordhoff E, Karas M, Cramer R, Hahner S, Hillenkamp F, et al. 1995. Direct mass spectrometric sequencing of low-picomole amounts of oligodeoxynucleotides with up to 21 bases by matrix-assisted laser desorption ionization mass spectrometry. *J. Mass Spectrom.* 30:99–112
73. Zhu L, Parr GR, Fitzgerald MC, Nelson CM, Smith LM. 1995. Oligodeoxynucleotide fragmentation in Maldi/Tof mass spectrometry using 355-Nm radiation. *J. Am. Chem. Soc.* 117:6048–56
74. Nordhoff E, Cramer R, Karas M, Hillenkamp F, Kirpekar F, et al. 1993. Ion stability of nucleic acids in infrared

matrix-assisted laser desorption/ionization mass spectrometry. *Nucleic Acids Res.* 21:3347–57

75. Little DP, Cornish TJ, O'Donnell MJ, Braun A, Cotter RJ, Koster H. 1997. MALDI on a chip: analysis of arrays of low femtomole to subfemtomole quantities of synthetic oligonucleotides and DNA diagnostic products dispensed by a piezoelectric pipet. *Anal. Chem.* 69:4540–46
76. Li J, Butler JM, Tan YP, Lin H, Royer S, et al. 1999. Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. *Electrophoresis* 20:1258–65
77. Monforte JA, Becker CH. 1997. High-throughput DNA analysis by time-of-flight mass spectrometry. *Nat. Med.* 3: 360–62
78. Taranenko NI, Allman SL, Golovlev VV, Taranenko NV, Isola NR, Chen CH. 1998. Sequencing DNA using mass spectrometry for ladder detection. *Nucleic Acids Res.* 26:2488–90
79. Tang W, Zhu L, Smith LM. 1997. Controlling DNA fragmentation in MALDI-MS by chemical modification. *Anal. Chem.* 69:302–12
80. Griffin TJ, Smith LM. 2000. Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. *Trends Biotechnol.* 18:77–84
81. Griffin TJ, Smith LM. 2000. Genetic identification by mass spectrometric analysis of single-nucleotide polymorphisms: ternary encoding of genotypes. *Anal. Chem.* 72:3298–302
82. Li J, Butler JM, Tan Y, Lin H, Royer S, et al. 1999. Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. *Electrophoresis* 20:1258–65
83. Fu DJ, Tang K, Braun A, Reuter D, Darnhofer-Demar B, et al. 1998. Sequencing exons 5 to 8 of the p53 gene by MALDI-TOF mass spectrometry. *Nat. Biotechnol.* 16:381–84
84. Tang K, Fu DJ, Julien D, Braun A, Cantor CR, Köster H. 1999. Chip-based genotyping by mass spectrometry. *Proc. Natl. Acad. Sci. USA* 96:10016–20
85. Griffin TJ, Hall JG, Prudent JR, Smith LM. 1999. Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry. *Proc. Natl. Acad. Sci. USA* 96:6301–6
86. Chen X, Fei Z, Smith LM, Bradbury EM, Majidi V. 1999. Stable-isotope-assisted MALDI-TOF mass spectrometry for accurate determination of nucleotide compositions of PCR products. *Anal. Chem.* 71:3118–25
87. Binnig G, Quate CF, Gerber C. 1986. Atomic force microscope. *Phys. Rev. Lett.* 56:930–33
88. Hirschfeld T. 1976. Optical microscopic observations of single small molecules. *Appl. Opt.* 15:2965–66
89. Shera EB, Seitzinger NK, Davis LM, Keller RA, Soper SA. 1990. Detection of single fluorescent molecules. *Chem. Phys. Lett.* 174:553–57
90. Hladky SB, Haydon DA. 1970. Discreteness of conductance change in bimolecular lipid membranes in the presence of certain antibiotics. *Nature* 5231:451–53
91. Hladky SB, Haydon DA. 1972. Ion transfer across lipid membranes in the presence of gramicidin A. I. Studies of the unit conductance channel. *Biochim. Biophys. Acta* 274:294–312
92. Neher E, Sakmann B. 1976. Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature* 260:799–802
93. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. 1996. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. USA* 93:13770–73
94. Woolley AT, Guillemette C, Li Cheung C, Housman DE, Lieber CM. 2000. Direct haplotyping of kilobase-size DNA using

carbon nanotube probes. *Nat. Biotechnol.* 18:760–63
95. Essevaz-Roulet B, Bockelmann U, Heslot F. 1997. Mechanical separation of the complementary strands of DNA. *Proc. Natl. Acad. Sci. USA* 94:11935–40
96. Bockelmann U, Essevaz-Roulet B, Heslot F. 1998. DNA strand separation studied by single molecule force measurements. *Phys. Rev. E* 58:2386–94
97. Jett JH, Keller RA, Martin JC, Marrone BL, Moyzis RK, et al. 1989. High-speed DNA sequencing: an approach based upon fluorescence detection of single molecules. *J. Biomol. Struct. Dyn.* 7:301–9
98. Voss D. 1999. Gene express. *New Sci.* 164:40–43
99. Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE. 1996. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* 274:1859–66
100. Akeson M, Branton D, Kasianowicz JJ, Brandin E, Deamer DW. 1999. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys. J.* 77:3227–33
101. Meller A, Nivon L, Brandin E, Golovchenko J, Branton D. 2000. Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl. Acad. Sci. USA* 97:1079–84
102. Akeson M, Deamer DW, Vercoutere W, Braslau R, Olsen H. 1999. *Use of a nanoscale pore to read short segments within single polynucleotide molecules*. Presented at NATO Adv. Res. Workshop Polymer Struct. Transp. Confined Spaces, Bikal, Hungary
103. Deamer DW, Akeson M. 2000. Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends Biotechnol.* 18:147–51
104. Vercoutere W, Akeson M, Olsen H, Deamer DW. 2000. Analysis of hairpin structures within single DNA molecules using a nanopore detector. *Biophys. J.* 78:402a
105. Alper J. 1999. From the bioweapons trenches, new tools for battling microbes. *Science* 39:1754–55
106. Bayley H. 1999. Designed membrane channels and pores. *Curr. Opin. Biotechnol.* 10:94–103
107. Baumeister B, Sakai N, Matile S. 2000. Giant artificial ion channels formed by self-assembled, cationic rigid-rod beta-barrels. *Angew. Chem. Int. Ed. Eng.* 39: 1955–55
108. Hansma HG, Hoh JH. 1994. Biomolecular imaging with the atomic force microscope. *Annu. Rev. Biophys. Biomol. Struct.* 23:115–39
109. Bustamante C, Rivetti C, Keller DJ. 1997. Scanning force microscopy under aqueous solutions. *Curr. Opin. Struct. Biol.* 7:709–16
110. Wong SS, Woolley AT, Odom TW, Huang JL, Kim P, et al. 1998. Single-walled carbon nanotube probes for high-resolution nanostructure imaging. *Appl. Phys. Lett.* 73:3465–67
111. Mou JX, Czajkowsky DM, Zhang YY, Shao ZF. 1995. High-resolution atomic-force microscopy of DNA—the pitch of the double helix. *FEBS Lett.* 371:279–82
112. Thess A, Lee R, Nikolaev P, Dai HJ, Petit P, et al. 1996. Crystalline ropes of metallic carbon nanotubes. *Science* 273:483–87
113. Murray MN, Hansma HG, Bezanilla M, Sano T, Ogletree DF, et al. 1993. Atomic force microscopy of biochemically tagged DNA. *Proc. Natl. Acad. Sci. USA* 90:3811–14
114. Sun LF, Xie SS, Liu W, Zhou WY, Liu ZQ, et al. 2000. Materials—creating the narrowest carbon nanotubes. *Nature* 403:384
115. Vettiger P, Despont M, Drechsler U, Durig U, Haberle W, et al. 2000. The "Millipede"—more than one thousand

tips for future AFM data storage. *IBM J. Res. Dev.* 44:323–40
116. Vettiger P, Brugger J, Despont M, Drechsler U, Durig U, et al. 1999. Ultrahigh density, high-data-rate NEMS-based AFM data storage system. *Microelectron. Eng.* 46:11–17
117. Minne SC, Yaralioglu G, Manalis SR, Adams JD, Zesch J, et al. 1998. Automated parallel high-speed atomic force microscopy. *Appl. Phys. Lett.* 72:2340–42
118. Fisher TE, Marszalek PE, Fernandez JM. 2000. Stretching single molecules into novel conformations using the atomic force microscope. *Nat. Struct. Biol.* 7: 719–24
119. Smith SB, Finzi L, Bustamante C. 1992. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science* 258:1122–26
120. Cluzel P, Lebrun A, Heller C, Lavery R, Viovy JL, et al. 1996. DNA—an extensible molecule. *Science* 271:792–94
121. Smith SB, Cui YJ, Bustamante C. 1996. Overstretching B-DNA—the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* 271:795–99
122. Wang MD, Yin H, Landick R, Gelles J, Block SM. 1997. Stretching DNA with optical tweezers. *Biophys. J.* 72:1335–46
123. Bustamante C, Smith SB, Liphardt J, Smith D. 2000. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* 10:279–85
124. Perkins TT, Smith DE, Larson RG, Chu S. 1995. Stretching of a single tethered polymer in a uniform flow. *Science* 268:83–87
125. Rief M, Clausen-Schaumann H, Gaub HE. 1999. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.* 6:346–49
126. Viovy JL, Heller C, Caron F, Cluzel P, Chatenay D. 1994. Ultrafast sequencing of DNA by mechanical opening of the double helix—a theoretical investigation. *C. R. Acad. Sci. Ser. Life Sci.* 317:795–800
127. Dapprich J. 1999. Single molecule DNA digestion by lambda exonuclease. *Cytometry* 36:163–68
128. Sauer M, Angerer B, Han KT, Zander C. 1999. Detection and identification of single dye labeled mononucleotide molecules released from an optical fiber in a microcapillary: first steps towards a new single molecule DNA sequencing technique. *Phys. Chem. Chem. Phys.* 1:2471–77
129. Ambrose WP, Goodwin PM, Jett JH, Van Orden A, Werner JH, Keller RA. 1999. Single molecule fluorescence spectroscopy at ambient temperature. *Chem. Rev.* 99:2929–56
130. Werner JH, Cai H, Goodwin PM, Keller RA. 1999. *Current status of DNA sequencing by single molecule detection*. Presented at SPIE Conf. Adv. Nucleic Acid Monit., Manip., Seq. Technol., San Jose, CA
131. Zander C, Sauer M, Drexhage KH, Ko DS, Schulz A, et al. 1996. Detection and characterization of single molecules in aqueous solution. *Appl. Phys. B* 63:517–23
132. Zander C. 2000. Single-molecule detection in solution: a new tool for analytical chemistry. *Fresenius J. Anal. Chem.* 366:745–51
133. Sauer M, Arden JJ, Drexhage KH, Gobel F, Lieberwirth U, et al. 1998. Time-resolved identification of individual mononucleotide molecules in aqueous solution with pulsed semiconductor lasers. *Bioimaging* 6:14–24
134. Van Orden A, Machara NP, Goodwin PM, Keller RA. 1998. Single molecule identification in flowing sample streams by fluorescence burst size and intraburst fluorescence decay rate. *Anal. Chem.* 70: 1444–51
135. Machara NP, Goodwin PM, Enderlein J,

Semin DJ, Keller RA. 1998. Efficient detection of single molecules eluting off an optically trapped microsphere. *Bioimaging* 6:33–42

136. Goodwin PM, Ambrose WP, Keller RA. 1996. Single-molecule detection in liquids by laser-induced fluorescence. *Acc. Chem. Res.* 29:607–13
137. Enderlein J, Goodwin PM, Van Orden A, Ambrose WP, Erdmann R, Keller RA. 1997. A maximum likelihood estimator to distinguish single molecules by their fluorescence decays. *Chem. Phys. Lett.* 270:464–70
138. Cantor CR, Smith C. 1999. *Genomics: The Science and Technology Behind the Human Genome Project*. New York: Wiley