Automated DNA Sizing by Atomic Force Microscopy

Thomas S. Spisz, Ye Fang, Isaac N. Bankman, Roger H. Reeves, and Jan H. Hoh

DNA sizing is one of the most widely used analytical methods in molecular biology and biochemistry, and it is a core method for genomic analysis. Many research, health care, and forensic applications rely on DNA sizing. Given the volume of DNA sizing done and the associated cost, an increasing need exists for extremely rapid, high-throughput, and inexpensive techniques. Essentially all DNA sizing today is performed by gel electrophoresis, although multiple efforts are under way to develop more effective technologies based on entirely new approaches. One such effort is described in this article: the use of atomic force microscopy in combination with pattern recognition software to perform DNA sizing. Initial results show that this approach can be compared with gel electrophoresis for some applications. Future work aims at developing a practical, high-throughput DNA sizing system. (Keywords: Atomic force microscopy, DNA sizing, Image processing.)

INTRODUCTION

Purpose of Sizing DNA

Why is DNA sizing important? Combined with other information, the size of a DNA fragment can contain information about the sequence within the fragment, or the DNA from which the fragment was derived. This information can be used in a wide range of applications, such as construction of physical genome maps and genotyping. Physical maps of DNA clones provide key information for completing the sequence of large genomes (e.g., the Human Genome Project). Physical map construction is also important for positional cloning, a process of isolating genes that have only been identified genetically. One type of physical map, called a restriction map, represents the locations along the DNA cut by restriction enzymes. Genotyping is the process of characterizing genetic differences throughout the genome among different people. These differences are used to follow inheritance of genes, to identify people having variants of genes (alleles) that predispose to specific diseases, and to clearly identify different people, such as in forensic applications.

The genomes of most organisms are too large to examine as a whole, so DNA is divided into manageable-sized fragments. For example, sequencing the entire human genome of approximately 3 billion base pairs (bp) requires that much smaller sequences (≈500 bp)
be determined at a time. The process of breaking the DNA into manageable pieces and sorting those pieces requires DNA sizing.

The information produced from DNA sizing mainly depends on how the fragments are generated, although biochemical techniques are used for most sizing applications. Some common methods are synthesis of complementary strands from a template DNA, cutting DNA at specific sites by restriction enzymes, or replicating a small portion of the DNA using a process called PCR (polymerase chain reaction).

**Approaches to DNA Sizing**

Today, gel electrophoresis methods are used frequently for DNA sizing. By combining the gel type and size with the applied electric field strength, gel electrophoresis can be used to size DNA molecules from a few bases to millions of bases. This method is relatively simple, the equipment needed is inexpensive, and the results are widely understood and accepted. Hence, for laboratory research, the approach is very useful. Gel electrophoresis has become extremely sophisticated in recent years, but in principle the current technology is the result of incremental improvements to the basic approach that is over 40 years old. Its primary limitations are the slow speed (running one gel often takes 2 h or more) and the need for large sample amounts of DNA (typically hundreds of nanograms)—both factors that directly affect cost. Thus, alternative approaches aim at increasing speed and reducing the amount of sample required.

Of the alternative approaches now emerging, the use of optical microscopy for genomic analysis is the most mature. Optical mapping uses light microscopy to determine the length of fluorescently stained DNA restriction fragments. This method has advantages over the gel-based method in throughput, resolution, safety, and cost. However, optical microscopy is not very effective for smaller DNA fragments; the resolution is diffraction limited at about 200 nm, which is approximately 600 bp.

Other approaches to DNA sizing include flow cytometry, mass spectrometry, electron microscopy, electrophysiology, and chip-based technologies. Flow cytometry, which sorts DNA molecules based on labeling with fluorescent markers, has been used extensively to size and separate larger DNA fragments such as chromosomes. At this time its lower limit has been improved to about 1500 bp, even larger than the lower limit for optical microscopy. Mass spectrometry is also emerging as a tool for genotyping and possibly DNA sizing. Today, however, it cannot easily be used on DNA molecules larger than 100 bp in length. Chip-based technologies have taken over certain applications for which DNA sizing by gel electrophoresis was previously used, but do not report fragment size. Electron microscopy is widely used for visualizing the structure of small objects, including DNA. However, the relatively cumbersome sample preparation and operation of the instrument have prevented its use in high-throughput applications. A recently proposed approach for DNA sizing is the use of ion channels and electrophysiology. In this approach, the occlusion of the channel by DNA as it passes through results in changes in current flow through the channel. This blockage of the channel is quantified using electrophysiological methods and is related to the length of the DNA.

The atomic force microscope obtains height images by scanning a surface with an extremely small tip at the end of a cantilever. The height is determined at positions along the scan using voltage measurements sampled from a split segment photodiode array aligned with a laser beam deflected off the top of the cantilever, as shown in Fig. 1. Optimally, the atomic force microscope can resolve single atoms (0.2-nm resolution). Under more typical imaging conditions the resolution is 2 to 20 nm, which is still more than an order of magnitude better than optical microscopy (>200 nm). The next section describes the merits of DNA sizing using atomic force microscopy (AFM).

**ATOMIC FORCE MICROSCOPY**

It is well established that AFM can easily visualize nucleic acids adsorbed to a surface, and that these images reflect the size of the DNA molecule. Other approaches to DNA sizing include flow cytometry, mass spectrometry, electron microscopy, electrophysiology, and chip-based technologies. Flow cytometry, which sorts DNA molecules based on labeling with fluorescent markers, has been used extensively to size and separate larger DNA fragments such as chromosomes. At this time its lower limit has been improved to about 1500 bp, even larger than the lower limit for optical microscopy. Mass spectrometry is also emerging as a tool for genotyping and possibly DNA sizing. Today, however, it cannot easily be used on DNA molecules larger than 100 bp in length. Chip-based technologies have taken over certain applications for which DNA sizing by gel electrophoresis was previously used, but do not report fragment size. Electron microscopy is widely used for visualizing the structure of small objects, including DNA. However, the relatively cumbersome sample preparation and operation of the instrument have prevented its use in high-throughput applications. A recently proposed approach for DNA sizing is the use of ion channels and electrophysiology. In this approach, the occlusion of the channel by DNA as it passes through results in changes in current flow through the channel. This blockage of the channel is quantified using electrophysiological methods and is related to the length of the DNA.

The atomic force microscope obtains height images by scanning a surface with an extremely small tip at the end of a cantilever. The height is determined at positions along the scan using voltage measurements sampled from a split segment photodiode array aligned with a laser beam deflected off the top of the cantilever, as shown in Fig. 1. Optimally, the atomic force microscope can resolve single atoms (0.2-nm resolution). Under more typical imaging conditions the resolution is 2 to 20 nm, which is still more than an order of magnitude better than optical microscopy (>200 nm). The next section describes the merits of DNA sizing using atomic force microscopy (AFM).

![Atomic Force Microscope Image](https://via.placeholder.com/150)

Figure 1. An atomic force microscope scans a surface with an extremely small probe tip at the end of a cantilever. Voltage measurements from a laser beam deflected off the top of the cantilever to a split segment photodiode are used to create height images.

1. T. S. Spietz ET AL
2. The information produced from DNA sizing mainly depends on how the fragments are generated, although biochemical techniques are used for most sizing applications. Some common methods are synthesis of complementary strands from a template DNA, cutting DNA at specific sites by restriction enzymes, or replicating a small portion of the DNA using a process called PCR (polymerase chain reaction).
3. Approaches to DNA Sizing
   - **Today, gel electrophoresis methods are used frequently for DNA sizing. By combining the gel type and size with the applied electric field strength, gel electrophoresis can be used to size DNA molecules from a few bases to millions of bases.** This method is relatively simple, the equipment needed is inexpensive, and the results are widely understood and accepted. Hence, for laboratory research, the approach is very useful. Gel electrophoresis has become extremely sophisticated in recent years, but in principle the current technology is the result of incremental improvements to the basic approach that is over 40 years old. Its primary limitations are the slow speed (running one gel often takes 2 h or more) and the need for large sample amounts of DNA (typically hundreds of nanograms)—both factors that directly affect cost. Thus, alternative approaches aim at increasing speed and reducing the amount of sample required.
   - Of the alternative approaches now emerging, the use of optical microscopy for genomic analysis is the most mature. Optical mapping uses light microscopy to determine the length of fluorescently stained DNA restriction fragments. This method has advantages over the gel-based method in throughput, resolution, safety, and cost. However, optical microscopy is not very effective for smaller DNA fragments; the resolution is diffraction limited at about 200 nm, which is approximately 600 bp.
   - Other approaches to DNA sizing include flow cytometry, mass spectrometry, electron microscopy, electrophysiology, and chip-based technologies. Flow cytometry, which sorts DNA molecules based on labeling with fluorescent markers, has been used extensively to size and separate larger DNA fragments such as chromosomes. At this time its lower limit has been improved to about 1500 bp, even larger than the lower limit for optical microscopy. Mass spectrometry is also emerging as a tool for genotyping and possibly DNA sizing. Today, however, it cannot easily be used on DNA molecules larger than 100 bp in length. Chip-based technologies have taken over certain applications for which DNA sizing by gel electrophoresis was previously used, but do not report fragment size. Electron microscopy is widely used for visualizing the structure of small objects, including DNA. However, the relatively cumbersome sample preparation and operation of the instrument have prevented its use in high-throughput applications. A recently proposed approach for DNA sizing is the use of ion channels and electrophysiology. In this approach, the occlusion of the channel by DNA as it passes through results in changes in current flow through the channel. This blockage of the channel is quantified using electrophysiological methods and is related to the length of the DNA.

4. **ATOMIC FORCE MICROSCOPY**
   - It is well established that AFM can easily visualize nucleic acids adsorbed to a surface, and that these images reflect the size of the DNA molecule. Other approaches to DNA sizing include flow cytometry, mass spectrometry, electron microscopy, electrophysiology, and chip-based technologies. Flow cytometry, which sorts DNA molecules based on labeling with fluorescent markers, has been used extensively to size and separate larger DNA fragments such as chromosomes. At this time its lower limit has been improved to about 1500 bp, even larger than the lower limit for optical microscopy. Mass spectrometry is also emerging as a tool for genotyping and possibly DNA sizing. Today, however, it cannot easily be used on DNA molecules larger than 100 bp in length. Chip-based technologies have taken over certain applications for which DNA sizing by gel electrophoresis was previously used, but do not report fragment size. Electron microscopy is widely used for visualizing the structure of small objects, including DNA. However, the relatively cumbersome sample preparation and operation of the instrument have prevented its use in high-throughput applications. A recently proposed approach for DNA sizing is the use of ion channels and electrophysiology. In this approach, the occlusion of the channel by DNA as it passes through results in changes in current flow through the channel. This blockage of the channel is quantified using electrophysiological methods and is related to the length of the DNA.

5. Figure 1. An atomic force microscope scans a surface with an extremely small probe tip at the end of a cantilever. Voltage measurements from a laser beam deflected off the top of the cantilever to a split segment photodiode are used to create height images.
to directly detect details of DNA such as supercoiled geometries,\textsuperscript{19} kinks,\textsuperscript{20} and varied helical repeats of single molecules,\textsuperscript{21} and to directly map the specificity and structure of DNA-binding proteins bound to molecules,\textsuperscript{22-25} thereby providing additional types of valuable information. These added capabilities can have considerable advantages over other approaches and are discussed later in this article.

General Approach

The present approach is founded on two premises: First, sizing of DNA will continue to be a major analytical method used in genomic analysis for the foreseeable future. Thus, the proposed work will not be "orphaned" by shifts in specific applications. Second, AFM currently has a higher signal-to-noise ratio than any other analytical technology that can be applied to nucleic acids. This has two primary consequences:

1. AFM can determine the size of a single piece of DNA whose length is near or below the optical diffraction limit.
2. AFM can directly detect small proteins and other markers bound to a single piece of DNA without contrast enhancement.

The general strategy proposed is shown in Fig. 3 as a comparison to gel electrophoresis. A DNA sample in solution is adsorbed to a solid support and imaged by AFM; the lengths of the DNA molecules in the sample are determined by automated image processing software, resulting in a histogram of sizes essentially identical to that produced by the scan of an electrophoresis gel. We call this approach solid-state DNA sizing (SSDS).\textsuperscript{26} The results indicate that SSDS clearly works in principle and on a limited scale in practice. As mentioned earlier, the core issue is implementing this approach and making it practical.

Automatic Sizing Software

To determine rapidly and accurately the distribution of DNA lengths in AFM images, we have developed specialized software written in the Interactive Data Language. This language is a high-level interactive programming environment designed to handle and manipulate large arrays of data and has many advanced processing tools. The automatic sizing software was developed to be platform independent and has a user-friendly graphical interface. The drawback of this language is that it is slower than typical compiled computer languages such as C/C++. However, the processing time per image is still less than 60 s, which allows more than 1500 images per day to be processed on an inexpensive personal computer.

The input to the software is the raw AFM image data of DNA adsorbed onto a substrate. The software

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{An AFM image of DNA on mica.}
\end{figure}

suitable for DNA smaller than 10,000 bp. We estimate the practical lower limit to be less than 50 bp, depending on the imaging resolution. At smaller sizes of DNA, the convolution of the image with the cantilever tip shape obscures the shape (and thus the length) of the molecule.

Hence, the question of whether AFM becomes a useful DNA sizing technology lies not in its ability to size, but rather whether a practical and useful implementation can be developed. This question of implementation is at the core of our present efforts. AFM has been used extensively for quality control of semiconductors. These systems use atomic force microscopes that run 24 h per day with automated sample exchange, tip positioning (to better than 1-µm accuracy), tip exchange, and data validation. Thus, many implementation issues for automating such a microscope have been solved. Others yet to be solved are sample deposition, high-resolution imaging in a noisy environment, and automated pattern recognition for DNA size determination.

In general terms, AFM falls into the category of single molecule technologies. One of the strongest trends in analytical chemistry in the past several decades is the miniaturization of technology, which is driven by considerations such as reduced sample size, reagent cost, and labor and analysis time. At the low end of the road to miniaturization are the single-molecule detection techniques, including AFM. Thus, AFM capabilities are consistent with current trends in analytical technology.

An additional advantage of AFM is the high signal-to-noise ratio, which allows direct visualization of individual DNA or protein molecules without contrast enhancing agents.\textsuperscript{14-18} Further, AFM has the potential...
A summary is presented here. The length of the DNA in these tests is given both in base pairs and nanometers, and the relative scale for these DNA is 1 bp equals 0.338 nm.

Accuracy is determined by testing a series of DNA samples with known size. Typically the measured length was within 15% of the expected length based on the known size. The precision was assessed by the width of the distribution peak at half height. This value was typically less than 10% of the fragment length for fragments larger than 200 bp. The sizes obtained by hand tracing the DNA in the images are similar to those obtained with the automated sizing program (Fig. 5). The hand tracing was accomplished by using NIH Image software and interactively tracing over the DNA fragment. Results from hand tracing can vary depending on the way in which the end points of the molecule are determined; the sizing software produces the same results each time.

The precision also gives one measurement of the resolution. A more practical assessment of resolution is provided by examining known mixtures of DNA fragments. Figure 6 shows the results from an experiment where a 200-bp fragment can be distinguished from a 250-bp fragment. In another experiment with a sample containing fragments from 100 to 2000 bp, the different fragments can be easily separated in the histogram of size distributions shown in Fig. 7. The longer DNA fragments in these tests are less well represented both in the histograms and in the images that were processed because the adsorption kinetics of longer fragments are relatively slow.

For a practical test application, a blind comparison with agarose gel electrophoresis for screening a set of P1 artificial chromosomes (PACs) was completed. A PAC library (BAC PAC Resources, Roswell Park Cancer Institute, Buffalo, New York) was screened to generate the PAC set examined. A single yeast artificial chromosome (YAC) of 350 kb was isolated from a pulsed field gel, radioactively labeled, and hybridized to PAC library filters. The resulting 20 positive clones were screened for the presence of known markers on the YAC using PAC DNA as templates for either 15-or 35-cycle PCR reactions. The same set of PCR products was examined by the AFM-based sizing method. The results from SSDS were in excellent agreement with the gel-based results (Fig. 8).

The error rate remains to be formally established, although we estimate the present error rate to be between 0.1 and 1%. Errors typically occur when the sample preparation fails in some way, so that no sample is present on the substrates. The time required for SSDS is several minutes per sample. For a small number of samples this technique is better than gel electrophoresis, which often requires hours to perform. However, gel electrophoresis allows for simultaneous processing of...
Figure 4. Examples of intermediates from the automated DNA sizing software. (a) AFM image. (b) After adaptive thresholding based on the optimum threshold obtained by Gaussian fitting of both foreground and background gray scale data, and followed by smoothing the binary image with a 3 x 3 pixel average. (c) After removing edge-touching fragments and very small artifacts. (d) After pruning and thinning by using a fast parallel thinning algorithm, the pixels are removed only if a set of conditions for each pixel's neighborhood is met, until no further pixels satisfy the conditions for removal. (e) After removing overlapping fragments, returning end-pixels that were deleted by thinning, and then calculating the length in nanometers of each remaining fragment. (f) Histogram of the calculated lengths.

Figure 5. Comparison of automated sizing to hand tracing of the same image of a sample containing fragments 250 bp long, showing the accuracy and consistency.

Figure 6. Empirical evaluation of the resolution using a sample of an equimolar mixture of 200- and 250-bp DNA, showing two resolved peaks near the actual sizes in the histogram.
The time per sample is similar for the two approaches. (In one set of experiments, one person completed manual processing of 60 PCR samples in about 4 h.) Many steps of this process are amenable to automation and should provide rapid throughput. The emergence of parallel tip arrays may further improve the throughput in SSDS.

Other Capabilities of AFM

The high signal-to-noise ratio of AFM and the resulting ability to detect proteins bound to the DNA, or other sequence-specific modifications, lead to significant potential advantages over gel electrophoresis. In particular, it now becomes possible to "decorate" the DNA in a way that increases the specific information obtained about the molecule. In fact, decorating DNA is analogous to bar coding because it specifies the identity of a piece of DNA with a certain level of confidence.

Decorating DNA is relevant to some sizing applications such as restriction mapping, but it is more powerful than restriction mapping as currently implemented. Restriction mapping locates sequence-specific positions along a piece of DNA that are cut by restriction enzymes. The procedure involves first sizing the DNA fragments resulting from a piece of DNA cut by enzymes. Then the order of the fragments along the original piece of DNA must be resolved, which is an inefficient and redundant process. The high contrast of AFM allows bound proteins, such as "stalled" restriction enzymes, to be visualized on the DNA, thereby allowing restriction maps to be determined directly by the distance between the bound proteins on the DNA. This ability to directly determine the sequence-specific decoration pattern eliminates the step required to resolve the order of the fragments.

Figure 7. Length distribution of common DNA reference sizes mixed in a single sample containing 100-, 200-, 400-, 800-, 1200-, and 2000-bp fragments, indicating the good sizing capability over a large range.

Figure 8. Comparison of AFM-based sizing with agarose gel electrophoresis for screening 370-bp DNA products. Results of the two methods were in excellent agreement. (a) Number of DNA fragments detected within a 30-nm window of the expected size for each of the 20 samples using AFM-based sizing. (b) Section of the gel along with a manually placed scan line across the DNA size to be screened (370 bp). (c) Plot of the intensity of the line scan indicating the relative number of DNA fragments in each of the 20 samples. (Reprinted from Ref. 26 by permission. © 1998 American Chemical Society.)
In addition, more types of sequence-specific decorations exist that can be used with AFM than just the restriction enzymes used for restriction mapping. Therefore, developing and implementing this decoration capability is also a high priority for our future efforts.

CONCLUSION

The conditions for AFM-based DNA sizing established here provide a basis for development of a highly automated, high-throughput assessment of DNA samples with a variety of applications in research, health care, and forensics. The potential for reduced time, smaller sample amounts, and lower reagent costs, as well as the available technologies for automation, suggest that AFM represents a competitive approach for DNA sizing and other types of DNA analyses.

REFERENCES


ACKNOWLEDGMENTS: This work was supported by NIH grants HG01518 (HHI, YT, and TSS), HG00405 (RHR), and internal research and development funds from The Johns Hopkins University Applied Physics Laboratory (TSS and INB). Others who have also contributed to this project are Neil D'Costa, Cheryl Seymour, and Tim Wiltshire.
YE FANG is a postdoctoral fellow in the Johns Hopkins School of Medicine's Department of Physiology. He earned a B.S. degree from Hunan University in 1989, an M.S. degree from Wuhan University in 1992, and a Ph.D. in biophysical chemistry from the Chinese Academy of Sciences in 1995. Dr. Fang's research interests include the relationship between DNA large-scale structure and microdomains, the biological application of atomic force microscopy, and the potential medical application of self-assembled nanosized drug carriers. His e-mail address is yangy@welchlink.welch.jhu.edu.

ISAAC N. BARKMAN received a B.S. degree in electrical engineering from Washington University in St. Louis, an M.S. degree in electronics from the University of Wales in England, and a Ph.D. in biomedical engineering from Technion University in Israel. He joined APL in 1990 and is a Senior Staff Engineer in the Electro-Optical Systems Group of the Air Defense Systems Department. He has worked on signal processing algorithms for target detection, image processing algorithms for recognition of small objects, and image registration algorithms. Dr. Barkman's current work includes analytical modeling of Ladar range Doppler signatures of ballistic missile parts, electro-optical systems for three-dimensional imaging with Ladar, laser vibrometry, and ATR algorithms for passive IR imaging. His e-mail address is isaac.barkman@jhuapl.edu.

ROGER H. REEVES is an Associate Professor of physiology at the Johns Hopkins School of Medicine. Dr. Reeves' research focuses on mouse and comparative genetics, protein analysis, animal models for genetic disease, Down's Syndrome, and Multiple Organ Dysfunction Syndrome. He is an associate editor of the Mammalian Genome, Journal of Heredity, Genomics, and Genome Research, and is also the director of the JHU SOM Transgenic Core Facility. His e-mail address is reeves@welchlink.welch.jhu.edu.

JAN H. HOH received a B.S. degree in biology and chemistry from Illinois State University in 1983 and a Ph.D. in cellular biology and biochemistry from the California Institute of Technology in 1990. In 1994, after his postdoctoral fellowship, he was appointed an Assistant Professor of physiology at the Johns Hopkins School of Medicine. Dr. Hoh's research interest focuses on biological and biomedical applications of scanned probe microscopy. He currently serves on the editorial board of the Biophysical Journal. His e-mail address is jhoh@jhu.edu.