LA-UR -91-562

Conf-911233--16

LA-UR--91-562

DE91 008591

TITLE

Rapid Sequencing of DNA Based on Single Molecule Detection

AUTHOR(S)

Steven A. Soper, Lloyd M. Davis, Frederic R. Fairfield, Mark L. Hammond, Carol A. Harger, James H. Jett, Richard A. Keller, Babbetta L. Marrone, John C.Martin, Harvey L. Nutter, E.Brooks Shera and Daniel J. Simpson

SUBMITTED TO

SPIE Proceedings

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

Rapid Sequencing of DNA Based On Single Molecule Detection

Steven A. Soper, Lloyd. M. Davis, Frederic R. Fairfield, Mark L. Hammond, Carol A. Harger, James H. Jett, Richard A. Keller, Babbetta L. Marrone, John C. Martin, Harvey L. Nutter, E. Brooks Shera, Daniel J. Simpson

Center for Human Genome Studies
Los Alamos National Laboratory
Los Alamos, NM 87545

1. ABSTRACT

Sequencing the human genome is a major undertaking considering the large number of nucleotides present in the genome and the slow methods currently available to perform the task. We have recently reported on a scheme to sequence DNA rapidly using a non-gel based technique. The concept is based upon the incorporation of fluorescently labeled nucleotides into a strand of DNA, isolation and manipulation of a labeled DNA fragment and the detection of single nucleotides using ultra-sensitive laser-induced fluorescence detection following their cleavage from the fragment. Detection of individual fluorophores in the liquid phase was accomplished with time-gated detection following pulsed-laser excitation. The photon bursts from individual rhodamine 6G (R6G) molecules travelling through a laser beam have been observed as have bursts from single fluorescently modified nucleotides. Using two different biotinylated nucleotides as a model system for fluorescently labeled nucleotides, we have observed synthesis of the complementary copy of M13 bacteriophage. Work with fluorescently labeled nucleotides is underway. We have observed and manipulated individual molecules of DNA attached to a microbead with an epifluorescence microscope.

2. INTRODUCTION

Presently, there is a major effort to map and sequence the human genome. This is a formidable task because the human genome contains 3×10^9 nucleotides. Currently used techniques can sequence a few hundred to a few thousands bases per day. The common sequencing protocols are those developed by Sanger (1) or Maxam and Gilbert (2) and are gel based techniques using either radioactively or fluorescently labeled nucleotides. Current methods require the use of one to four lanes of the gel and a vast number of identical DNA molecules which yield a few hundred bases of sequence (3,4). Longer DNA sequences are constructed by overlapping the short sequences. If the DNA sequence of interest is a million bases long, current methodologies of overlapping short sequences becomes prohibitive. While gel based sequencing techniques are improving, bases sequenced in a gel above about 1000 nucleotides is difficult and extensive

manpower and time are required for these gel techniques.

Recently, we have reported on a new method of sequencing DNA at a rate approaching several hundred bases per second (5,6). The technique involves: (a) labeling the nucleotides with base specific tags suitable for ultra-sensitive fluorescence detection, (b) enzymatic synthesis of a complementary strand of DNA using fluorescently-labeled nucleotides, (c) isolation and manipulation of a single molecule of fluorescently labeled DNA, (d) suspension of the single DNA molecule in a flowing sample stream, (e) sequential cleavage of fluorescently labeled nucleotides from the DNA and (f) detection of single fluorescently labeled nucleotides as they pass through a focused laser beam. The sequencing rate of this method should be limited by the rate at which the exonuclease can remove single DNA bases from the terminus of the DNA molecule and the rate of detection of single molecules. Our method should be able to determine the sequence of very long pieces of DNA (e.g. the 40 Kb DNA fragments in a cosmid library) directly without the need for overlapping short DNA sequences.

The success of this proposed method depends upon the ability to detect individual fluorescent molecules in solution as they transit a focused laser beam. Work in the area of single molecule detection (SMD) was initiated by Hirschfeld who labeled one molecule (polyethyleneimine) with fluorescein isothiocyanate molecules and was able to detect 80 fluorophores in a static system (7). Dovichi and coworkers utilized rhodamine 6G (R6G) as the fluorophore and hydrodynamic focused flows and were able to see a few thousand molecules (8,9). This work was followed by a report of sensitive fluorescence detection from molecules of phycoerythrin, equivalent to 25-30 R6G molecules based upon differences in the molar absorptivities and fluorescent quantum efficiencies (10). Peck et al. later demonstrated indirect proof of detection of single molecules of phycoerythrin in solution (11). The work with phycoerythrin was followed by improvements in sensitivity, approaching the single molecule level for the fluorophore R6G using CW excitation (12,13). We have recently reported on the first direct observation of the photon burst from individual R6G molecules travelling through a focused laser beam using pulsed-laser excitation and time-gated detection (14). The use of time-gated detection effectively discriminates the scattering background from the fluorescence and results in a substantial decrease in the observed background. Our recent progress in the area of single molecule detection will be discussed as well as our progress in the area of DNA replication using labeled nucleotides and the isolation and manipulation of individual molecules of DNA.

3. EXPERIMENTAL

The pulsed-laser SMD apparatus has been described elsewhere (14). Briefly, the excitation source was an actively mode-locked Nd:YAG laser with a repetition rate of 82 MHz and pulse width of 70 psec. The fundamental was frequency-doubled to 532 nm with average powers of 30 mW at the flow cell. Since the fluorescent lifetime and the inverse of the laser repetition rate are much shorter than the time the molecule spends in the laser beam, the molecule is re-excited many times resulting in a burst of photons that serves as a signature for the passage of

molecules through the laser beam. A microscope objective and a slit are arranged to image the photons from a small region around the laser beam waist onto a microchannel plate photomultiplier (MCP) operated in the single-photon counting mode (see Figure 1).

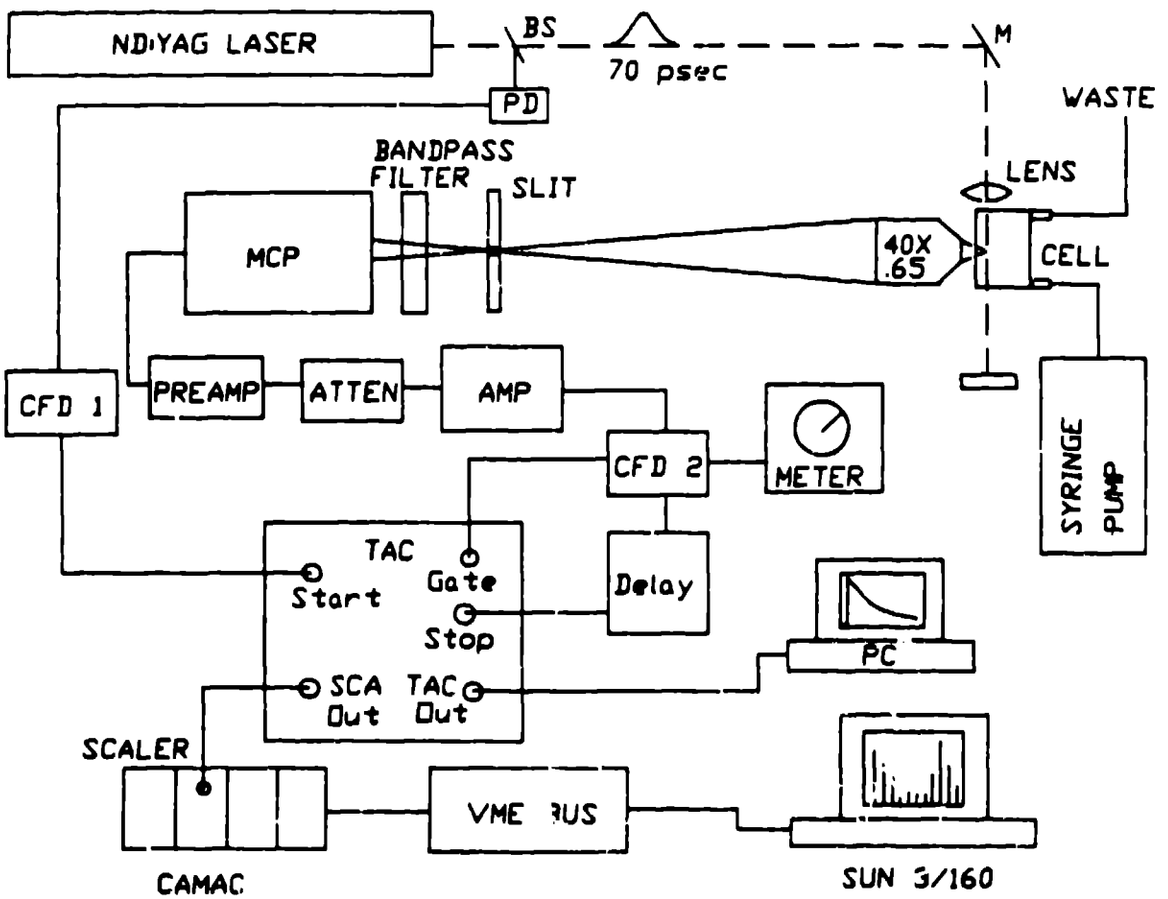

Figure 1. Schematic drawing of the SMD apparatus. The pulsed light was focused onto the 10x4 mm flow cell with a 17 mm focal length lens yielding a measured beam waist of 7.5 um (1/e). Part of the excitation beam was directed to a photodiode to provide the start pulse for the TAC. The stop pulse was generated from the anode pulses of the MCP PMT. The fluorescence emission was collected by a 40X, NA 0.65 microscope objective and imaged onto a vertical slit with a width set at 0.4 mm resulting in a 10 um observation distance along the propagation axis of the laser beam. Scattering impinging onto the MCP PMT was minimized with a bandpass filter centered at 580 nm and a FWHM of 40 nm

Photostabilities of various fluorophores were measured using the CW excitation from an Ar ion laser (514.5 nm). The remainder of the CW apparatus has been described in detail elsewhere (13). The photobleaching efficiencies were measured according to the procedure described by Mathies and Stryer (15). The method involves measuring the normalized fluorescence intensity as a function of the flow velocity, yielding a sigmoidally shaped curve from which the photobleaching efficiency can be obtained.

The DNA fragments were observed under a conventional epifluorescence microscope (Leitz Laborlux) equipped with a cooled Photometrics CCD camera detector for sensitive fluorescence detection. The DNAs were stained with ethidium bromide in order to observe the fluorescence of the DNA molecules.

4. SINGLE MOLECULE DETECTION OF LABELED NUCLEOTIDES

A number of different photophysical parameters play important roles in determining the ability to detect selected fluorophores on a single molecule level. The photobleaching efficiency sets an upper limit on the number of photons one can obtain per molecule and therefore plays a crucial role in determining the duration of the photon burst for the molecule under observation. Table 1 presents the fluorescence quantum efficiency (Φ_f), photobleaching efficiency (Φ_d) and the total number of photons attainable per molecule (Φ_f / Φ_d) for R6G, tetramethylrhodamine isothiocyanate (TRITC) and adenine labeled with TRITC (TRITC-AD). TRITC and TRITC-AD are typical fluorophores that will be used in the rapid sequencing scheme. In the case of R6G, approximately 25000 photons per molecule can be obtained in an aqueous solvent. For 0.001 photoelectrons

Table 1. The fluorescence quantum efficiency (Φ_f), photobleaching efficiency (Φ_d) and photon yield per molecule (N) for R6G, TRITC and TRITC-AD in H_2O and EtOH.

Solvent	R6G			TRITC			TRITC-AD		
	Φ_f	Φ_d	N	Φ_f	Φ_d	N	Φ_f	Φ_d	N
H_2O	0.95	1.8×10^{-5}	2.5×10^{4}	0.15	5.0×10^{-6}	3.0×10^{4}	0.14	6.4×10^{-6}	2.2×10^{4}
EtOH	0.98	6.0×10^{-7}	1.6×10^{6}	0.26	4.2×10^{-7}	6.2×10^{5}	0.22	6.4×10^{-7}	3.4×10^{5}

per photon generated (taking into account quantum efficiency of phototube, geometric collection efficiency and transmission efficiency of the filters), then approximately 25 photoelectrons are detected per molecule. In EtOH, the photon yield per molecule is 1.6×10^6 due to an 100 fold improvement in its photostability. We have been able to utilize the increased photon yield of R6G in EtOH to observe the bursts of photons from individual molecules of R6G using CW excitation as indicated from non-random correlations in the autocorrelation function and tails in the Poisson distributions (13). For TRITC and TRITC-AD, the fluorescence quantum yields are approximately 3X smaller than that of R6G in H_2O. But due to their increased photostability, these fluorophores result in nearly the same number of photons per molecule as that seen for R6G.

We are able to detect individual molecules of R6G transiting a focused laser beam using pulsed-laser excitation and time-gated detection (14). Single molecule detection was based on (a) the observation of a non-random correlation in the autocorrelation function and (b) the direct observation of the burst of photons from single R6G molecules in solution. The autocorrelation function, $G(\tau)$, for discretely sampled data can be expressed as

$$G(\tau) = \sum_{t=0}^{N-1} d(t)\, d(t+\tau), \qquad (1)$$

where N is the number of data points analyzed and τ is the delay. As can be seen from equation (1), the autocorrelation is performed on the entire data set and is thus computed over a large number of events. The non-random correlation is evidence for observing the bursts from a number of molecules passing through the laser beam during the course of the experiment and persist for delays up to the average residence time of the molecule within the laser beam. With the knowledge of the photobleaching rate, the flow velocity and the diameter of the laser beam, one can calculate the effective residence time of a molecule within the laser beam. In the case of TRITC or TRITC-AD in H_2O and laser powers of 30 mW the average effective lifetime of the fluorophore (before bleaching) is approximately 15 msec whereas the transit time (laser beam diameter / flow velocity) is approximately 30 msec. Therefore, a majority of the molecules are photobleached before exiting the laser beam. The autocorrelation function for 100 fM of TRITC-AD and for the water solvent are shown in Figure 2. A strong non-random autocorrelation was seen only in the case of TRITC-AD. This concentration of TRITC-AD was chosen to yield a probability of a single molecule residing within the laser beam at any given time of 0.1 thereby minimizing the probability of two molecules residing within the laser beam.

The autocorrelation is computed over a large number of events and does not identify the passage of individual molecules as they transit the laser beam, an essential requirement in the rapid sequencing methodology since each molecule

must be processed individually. In order to examine the data for passage of individual molecules, we have defined a weighted quadratic summing (WQS) filter given by (ref. 14)

$$S(t) = \sum_{\tau=0}^{k-1} w(\tau)\ d(t+\tau)^2, \qquad (2)$$

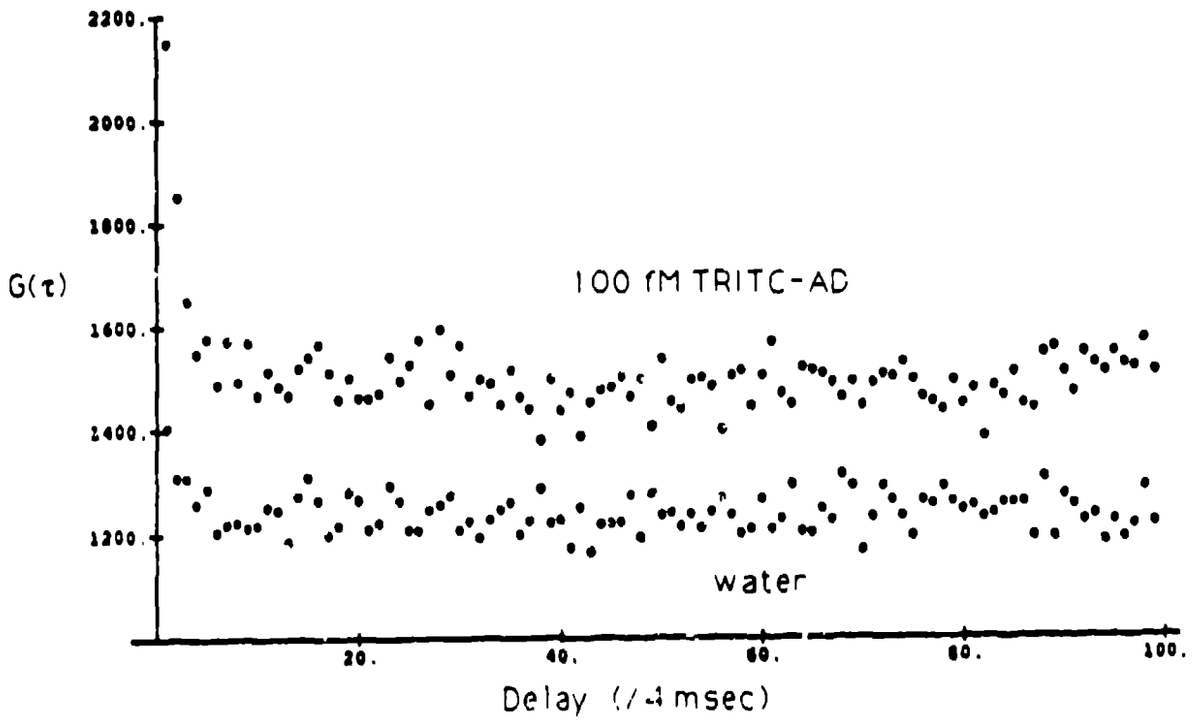

Figure 7. Autocorrelation plots for water and water with the addition of 100 fM of TRITC-AD. The average laser power was 30 mW and the autocorrelation was computed over 1.2 sec of data collection.

where k covers the time interval on the order of the molecular passage time (in the present experiment k = 5, corresponding to 20 msec, 4 msec per counting interval) and $w(r)$ are weighting factors chosen to best discriminate the signal due to passing molecules from random fluctuations in the background. The values of these weights were based on results from computer simulations of the expected signal from passage of individual molecules through the laser beam. In the present case $w(r) = (r + 1) / k$, for $r = 0$ to $k - 1$ (an asymmetric triangular ramp due to the fact that the signal increases slowly as the molecule enters the laser beam followed by an abrupt cessation of photon emission due to photobleaching). Figure 3 shows the WQS filtered data for the blank and 100 fM of TRITC-AD. Small amplitude bursts are observed in the blank due to statistical fluctuations of the background and to fluorescent impurities. Upon addition of TRITC-AD, large amplitude bursts are observed in the data. The average number of photoelectrons observed per burst is roughly 10, in accordance with the data of Table 1 (22000 photons per molecule) and the conversion efficiency of the pulsed-laser SMD apparatus (0.0007 photoelectrons / photon). Based upon the estimated flow velocity, the concentration of the fluorophore used in this experiment and the size of the observation volume (1.8 pL), the calculated number of molecules passing through the laser beam is approximately 1 per sec. If one sets a discriminator at $S(t) = 10$, then the associated detection efficiency for single molecules of TRITC-AD transiting the laser beam is nearly 70% with an error rate (due to fluorescence impurities present in the solvent blank and statistical fluctuations in the background) of approximately 0.03 per sec.

Our ability to detect individual molecules of the nucleotide adenine labeled with TRITC is significant not only in terms of our rapid DNA sequencing scheme, but for applying SMD to various types of analytical applications where fluorophores are attached to analytes. TRITC-AD shows a reduced quantum yield for fluorescence as compared to R6G (see Table 1), but, because of its increased photostability, results in similar photon yields per molecule. The results from reference 14 for R6G and that from Figure 3 indicate that selection of molecules for SMD should not be based solely on the fluorescence quantum yields, but should include considerations based upon photostability as well.

5. REPLICATION OF DNA WITH MODIFIED NUCLEOTIDES

In parallel with single molecule detection, we are also making progress with the biochemical methods necessary to label a fragment of DNA in preparation for sequencing. DNA will be labeled by reacting a single-stranded template with modified nucleotides in the presence of a polymerase enzyme. This will create a labeled, double-stranded DNA fragment. The current method for single molecule detection in the rapid DNA sequencing project requires that the nucleotides be labeled with a fluorescent tag due to the small quantum yields for fluorescence of the native nucleotides. Attachment of an appropriate fluorescent dye will be made via a linker arm to a position on the base. Because fluorescently tagged nucleotides were not yet commercially available, initial experiments utilized biotin-modified nucleotides to investigate the enzymatic synthesis of labeled DNA fragments and their subsequent cleavage by exonucleases. These experiments

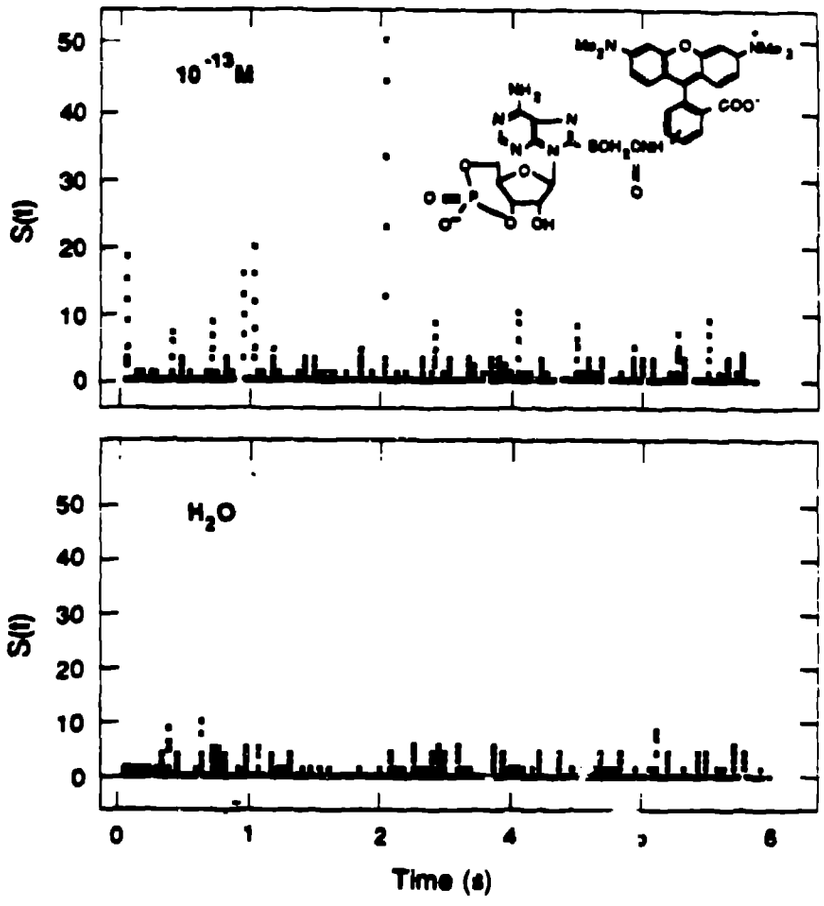

Figure 3. WQS filter plot for 100 fM of TRITC-AD in water and water with no added TRITC-AD. The structure of TRITC-AD is given in the upper right hand corner of the figure. At the concentration used in this experiment, approximately 1 molecule passes through the detection volume per second. The conditions of this experiment were the same as those of Figure 2.

demonstrated incorporation of biotin-labeled nucleotides into strands complementary to simple poly(dA, dG) DNA polymers as well as the exonucleolytic digestion of this biotin-labeled duplex (3). We have also synthesized strands complementary to more complex M13 constructs with two biotin-labeled nucleotides,

biotin-7-dATP and biotin-11-dUTP. Recently, fluorescently modified nucleotides have become available. Preliminary experiments suggest that these nucleotides incorporate poorly under standard reaction conditions. In some cases the rate of incorporation of unmodified nucleotides was inhibited by the fluorescent nucleotides. We anticipate that our current investigations into the mechanisms involved in this inhibition of DNA synthesis will allow us to determine appropriate linker arm structures and attachment positions so that we may design fluorescently tagged nucleotides that will incorporate rapidly and efficiently.

6. ISOLATION AND MANIPULATION OF DNA

Our technique requires the ability to select, attach and manipulate individual DNA molecules. To this end, we have been exploring the attachment of individual DNA molecules to supportive structures and the manipulation of these supported DNA molecules. As in the synthesis of modified DNA, we are using a model system to simulate the final tagged DNAs. The model system consists of bacteriophage lambda double-stranded DNA to mimic the size of our eventual tagged DNA (40 Kb) and ethidium bromide (a fluorescent dye that intercalates into DNA) to mimic the fluorescent tags. Using the cooled CCD camera coupled to a fluorescence microscope and appropriate filters to detect the fluorescence of ethidium stained DNA, we are able to observe small fluorescent objects that have the proper mobility, size and sensitivity to DNase expected of individual lambda DNA molecules. Performing these experiments was made feasible by the construction of a microscopic gel electrophoresis apparatus. The apparatus allows one to observe the motion, fluorescence intensity and digestion of the DNA while within the field of view of the microscope in order to confirm that the object under observation is, in fact, an individual molecule of lambda DNA. While we believe that we are isolating individual DNA molecules, there is a potential problem because lambda DNA can form aggregates under our solution and dilution conditions. We are pursuing electronic, enzymatic, and biological techniques to confirm that our smallest objects are indeed single molecules and not aggregates of a few molecules.

To manipulate individual molecules of DNA, some type of solid support is necessary. Our original contention was to use avidin-coated inert microbeads of a few microns in diameter (5). The avidin would bind biotin-modified nucleotides that had been previously incorporated into the modified DNA. Since limiting the number of attached DNAs to one by this method is technically difficult, we have attempted to support the DNA by its known ability to bind to glass microbeads. Even though this is also difficult, attaching apparently single DNA molecules to glass beads has been successful.

To move these supported DNAs into the sequencer, a method is needed to both transport the microsphere into the flow region and to hold it in place during the sequencing. We are currently evaluating optical traps (counter propagating focused laser beams that act as small tweezers (16)) and a number of different biological methods to make the manipulation process less tedious. Since supported DNAs must be digested one nucleotide at a time, we are also

investigating whether the presence of the support alters the enzymatic properties of the DNA exonucleases to create new products of the digestion or to make parts of the DNA unavailable for digestion.

7. CONCLUSIONS

The ability to sequence DNA based upon single molecule detection will have important ramifications in molecular biology. Although much work needs to be accomplished, significant progress has been made. We have successfully observed the individual photon bursts from nucleotides tagged with fluorogenic tags, completely replicated the bacteriophage M13 using two different biotinylated nucleotides, attached a single molecule of lambda DNA to a solid support and observed the fluorescence of this single DNA molecule under an epifluorescence microscope. Research will be focused on linker arm design to facilitate incorporation of fluorescently labeled nucleotides into nascent DNA, to suspend a supported strand of DNA in a flowing sample stream and expanding our single molecule detection capabilities to observe fluorophores of different colors in a single experiment.

8. ACKNOWLEDGMENTS

This research was supported by a grant from the Department of Energy, Office of Health and Environmental Research.

9. REFERENCES

1. F. Sanger, S. Nicklen and A.R. Coulson, Proc. Natl. Acad. Sci. USA, 74, 5463, (1977).

2. A.M. Maxam and W. Gilbert, Meth. Enzym., 65,4499(1980).

3. L.M. Smith, J.Z. Sanders, R.J. Kaiser, P. Hughes, C. Dodd, C.R. Cornell, C. Heiner, S.B.H. Kent and L.E. Hood, Nature, 321, 674, (1986).

4. J.M. Prober, G.L. Trainor, R.J. Dam, F.W. Hobbs, C.W. Robertson, R.J. Zagursky, A.J. Cocuzza, M.A. Jensen and K. Baumeister, Science, 238, 336, (1987).

5. J.H. Jett, R.A. Keller, J.C. Martin, B.L. Marrone, R.K. Moyzis, R.L. Ratliff, N.K. Seitzinger E.B. Shera and C.C. Stewart, J. Biomol. Struc. and Dynamics, 7, 301, (1989).

6. L.M. Davis, F.R. Fairfield, C.A. Harger. J.H. Jett, R.A. Keller, J.H. Hahn, L.A. Krakowski, B.L. Marrone, J.C. Martin, H.L. Nutter, R.L. Ratliff, E.B. Shera, D.J. Simpson and S.A. Soper, Genetic Anal. (in press).

7. T. Hirschfeld, Appl. Opt., 15, 2965, (1976).

8. N.J. Dovichi, J.C. Martin, J.H. Jett, M. Trukula and R.A. Keller, Anal. Chem., 56, 348, (1984).

9. N.J. Dovichi, J.C. Martin, J.H. Jett and R.A. Keller, Science, 219, 845, (1983).

10. D.C. Nguyen, R.A. Keller, J.H. Jett and J.C. Martin, Anal. Chem., 59, 2158, (1987).

11. K. Peck, L. Stryer, A.N. Glazer and R.A. Mathies, Proc. Natl. Acad. Sci. USA, 86, 4087, (1989).

12 H.H. Hahn, S.A. Soper, J.H. Jett, J.C. Martin, H.L. Nutter and R.A. Keller, Appl. Spectrosc. (accepted for publication).

13. S.A. Soper, E.B. Shera, J.C. Martin, J.H. Jett, J.H. Hahn, H.L. Nutter and R.A. Keller, Anal. Chem. (accepted for publication).

14. E.B. Shera, N.K. Seitzinger, L.M. Davis, R.A. Keller and S.A. Soper, Chem. Phys. Lett., 174, 553, (1990).

15. R.A. Mathies, A.R. Oseroff and L. Stryer, Proc. Natl. Acad. Sci. USA, 73, 1, (1976).

16. T.N. Buican, M.J. Smyth, H.A. Crissman, G.C. Salzman, C.C. Stewart and J.C. Martin, Appl. Opts., 26, 5311, (1987).