CHAPTER TWENTY

Real-Time DNA Sequencing from Single Polymerase Molecules

Jonas Korlach, Keith P. Bjornson, Bidhan P. Chaudhuri, Ronald L. Cicero, Benjamin A. Flusberg, Jeremy J. Gray, David Holden, Ravi Saxena, Jeffrey Wegener, *and* Stephen W. Turner

Contents

Abstract

Pacific Biosciences has developed a method for real-time sequencing of single DNA molecules (Eid *et al.*, 2009), with intrinsic sequencing rates of several bases per second and read lengths into the kilobase range. Conceptually, this sequencing approach is based on eavesdropping on the activity of DNA polymerase carrying out template-directed DNA polymerization. Performed in a

Pacific Biosciences, Menlo Park, California, USA

Methods in Enzymology, Volume 472
ISSN 0076-6879, DOI: 10.1016/S0076-6879(10)72001-2

highly parallel operational mode, sequential base additions catalyzed by each polymerase are detected with terminal phosphate-linked, fluorescence-labeled nucleotides. This chapter will first outline the principle of this single-molecule, real-time (SMRT™) DNA sequencing method, followed by descriptions of its underlying components and typical sequencing run conditions. Two examples are provided which illustrate that, in addition to the DNA sequence, the dynamics of DNA polymerization from each enzyme molecules is directly accessible: the determination of base-specific kinetic parameters from single-molecule sequencing reads, and the characterization of DNA synthesis rate heterogeneities.

1. Introduction

The ability to rapidly determine nucleic acid sequences has fundamentally transformed the biological sciences, both with respect to inquiries toward understanding biological processes and the approaches to manipulating them. Next-generation DNA sequencing methods have changed whole-genome sequencing projects into routine procedures (reviewed in Mardis, 2008) and have been adapted to other areas, such as transcriptome sequencing and epigenetics (Cloonan *et al.*, 2008; Cokus *et al.*, 2008; Fullwood *et al.*, 2009; Maher *et al.*, 2009; Yassour *et al.*, 2009). However, despite their gains in sequencing throughput, these methods still fall short of providing the means to elicit fundamental changes in the fields of medical diagnostics, disease prevention, and treatment. Further improvements are required for higher quality and even more cost-effective sequencing of complete individual genomes and transcriptomes.

DNA polymerases can be viewed as efficient DNA sequencers—engineered by nature—as they decode the sequence of a template strand by virtue of synthesizing its complementary strand. Over millions of years of molecular evolution, DNA polymerases have been optimized to rapidly and faithfully replicate genomes, and they have in turn developed many features attractive for artificial DNA sequencing methods. DNA polymerases can be very fast, with DNA synthesis rates reported *in vitro* as high as 750 bases per second (Tabor *et al.*, 1987). Tens to hundreds of thousands of bases can be synthesized from a single polymerase binding event (Blanco *et al.*, 1989). DNA polymerases can also be viewed as very frugal as only one nucleotide is consumed during each incorporation cycle. Error rates can be as low as one in 10^5 bases (Esteban *et al.*, 1993) and even lower with associated proofreading activities. Finally, DNA polymerases are physically very small, enabling a high level of multiplexing on a small footprint.

Exploiting all of these characteristics directly by using polymerase as the actual sequencing engine had not been commercially feasible until recently. We have overcome the underlying technical challenges by innovations in the fields of nanofabrication, surface derivatizations, nucleotide and protein chemistries, and optics, to enable the direct, real-time interrogation of individual polymerase activities (Eid *et al.*, 2009). Essentially, DNA is sequenced by watching with base-pair resolution what normally constitutes DNA replication occurring in dividing cells.

2. Principle of Single-Molecule, Real-Time DNA Sequencing

The Single-Molecule, Real-Time (SMRT^{TM}) DNA sequencing concept is illustrated in Fig. 20.1. The two principal technological components that facilitate SMRT sequencing are

(i) zero-mode waveguide (ZMW) confinement that allows single-molecule detection at concentrations of labeled nucleotides relevant to the enzyme, and

(ii) fluorescence-labeled, phospholinked nucleotides that permit observation of uninterrupted DNA polymerization.

ZMW nanostructures (Fig. 20.1A) consist of dense arrays of holes, ~100 nm in diameter, fabricated in a ~100-nm metal film deposited on a transparent substrate (e.g., silicon dioxide) (Foquet *et al.*, 2008; Levene *et al.*, 2003). Each ZMW becomes a nanophotonic visualization chamber for recording an individual polymerization reaction, providing a detection volume of just ~100 zeptoliters (10^{-21} L). This volume represents a ~1000-fold improvement over diffraction-limited confocal microscopy, making it possible to observe single nucleotide incorporation events against the background created by diffusing fluorescence-labeled nucleotides. In addition to reducing the number of labeled nucleotides present inside the observation volume, the highly confined volume results in drastically shorter diffusional visitation times. This enables better temporal differentiation between events involving diffusion of labeled nucleotides through the observation volume (now typically lasting only a few microseconds) and enzymatic nucleotide incorporation events (typically lasting several milliseconds for polymerases).

The second important component is phospholinked nucleotides for which the fluorescent label is attached to the terminal phosphate rather than the base, typically via a linker (Fig. 20.1B). A 100% replacement of unmodified nucleotides by phospholinked nucleotides is achieved because the enzyme cleaves away the fluorophore as part of the incorporation process, leaving behind a completely natural, double-stranded nucleic acid

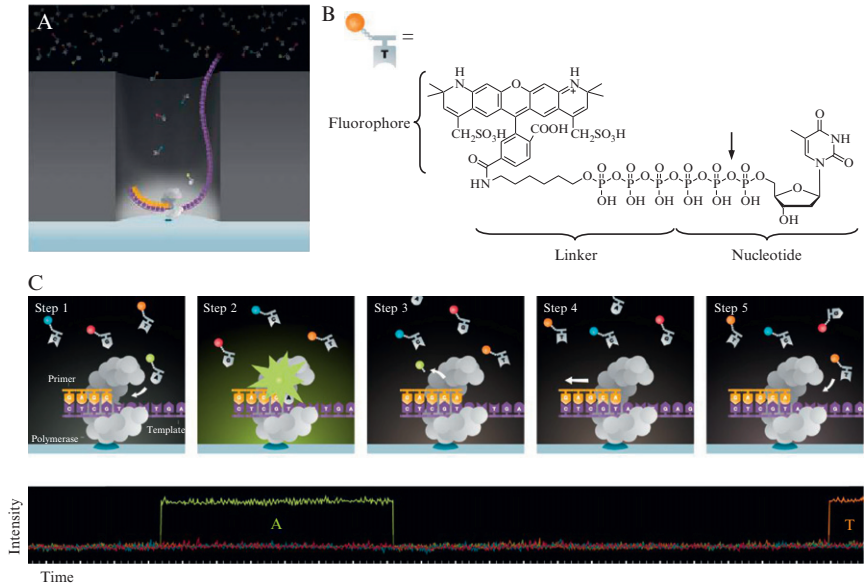

Figure 20.1 Principle of single-molecule, real-time (SMRT) DNA sequencing. (A) Single DNA polymerase molecules with bound DNA template are immobilized to the bottom of zero-mode waveguide (ZMW) nanostructure arrays. Polymerization of the complementary DNA strand is observed in real time by detecting enzymatic processing of fluorescent phospholinked nucleotides. (B) Molecular structure of phospholinked nucleotides. Alexa Fluor 568-aminohexyltriphosphate-dTTP is shown by example (Eid *et al.*, 2009). The arrow indicates the α-β phosphodiester bond cleavage mediated by the DNA polymerase. (C) Schematics of reactions steps involved in SMRT DNA sequencing (top), and corresponding fluorescence intensity time trace (bottom). Step 1: The DNA template/primer/polymerase complex is surrounded by diffusing phospholinked nucleotides which probe the active site. Step 2: A labeled nucleotide makes a cognate binding interaction with the template base in the DNA. During the time it is bound in the active site (typically lasting tens of milliseconds) fluorescence is emitted continuously, giving rise to a detectable pulse in the fluorescence intensity time trace. The identity of the fluorescent dye indicates which base is incorporated. Step 3: The polymerase incorporates the nucleotide into the growing nucleic acid chain by cleaving the α-β phosphodiester bond, thereby subsequently releasing the pyrophosphate-linker-fluorophore. Step 4: The polymerase translocates to the next template position. Step 5: The process repeats.

product. In SMRT sequencing, each of the four different nucleobases is labeled with a distinct fluorophore to discriminate base identities during incorporation events, thus providing sequence determination of the complementary DNA template (Fig. 20.1C). During incorporation, the enzyme holds the labeled nucleotide in the ZMWs detection volume for several milliseconds, orders of magnitude longer than the average diffusing nucleotide is present. Fluorescence is emitted continuously from the fluorophore

label during the incorporation process, causing a detectable pulse of increased fluorescence in the corresponding color channel. The pulse is terminated naturally by the polymerase releasing the pyrophosphate-linker-fluorophore group which diffuses out of the observation volume. The polymerase then translocates to the next base, and the process repeats.

3. Components of SMRT Sequencing

3.1. Zero-mode waveguides for observation volume confinement

Fabrication of ZMWs with aluminum or gold as the metal cladding material was first described using a positive-tone, electron-beam lithography technique followed by reactive ion etching (Levene *et al.*, 2003; Liu and Blair, 2003). Subsequently, other fabrication methods have been described, including negative-tone, electron-beam lithography followed by metallization and resist removal (Foquet *et al.*, 2008; Miyake *et al.*, 2008), focused ion beam milling (Rigneault *et al.*, 2005), and photolithography (Foquet *et al.*, 2008). Of these, photolithography is very attractive because of the lower cost of fabrication and its compatibility with high-volume manufacturing processes. We have recently improved the photolithographic process of ZMW fabrication, resulting in greater reproducibility, size uniformity, and ZMW shape control (Fig. 20.2). Detailed protocols for the different fabrication methods can be found in the references cited above.

3.2. ZMW surface derivatization for targeted enzyme immobilization

The selective placement of an active polymerase molecule into a ZMW observation volume, immediately above the transparent ZMW floor, is an important prerequisite for efficient SMRT DNA sequencing. ZMWs put additional demands on the quality of surface preparations to achieve this because the functionalization target area is very small relative to the cladding surface area. At the same time, the surfaces should be well-passivated to prevent corrosion and nonspecific adsorption of phospholinked nucleotides which are used at much higher concentrations compared to conventional single-molecule assays. In addition, the employed surface coating reagents should exhibit low fluorescence levels, and should not interfere with enzymatic activities.

For aluminum-clad ZMWs, we have developed protocols and reagent formulations meeting all of these criteria by exploiting an inherent feature of ZMW architecture in that the ZMW substrate and cladding are made of

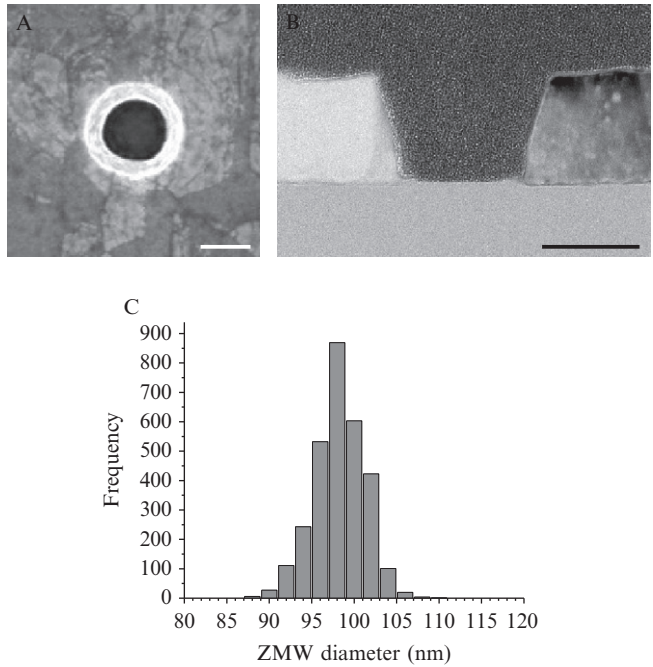

Figure 20.2 Zero-mode waveguide (ZMW) fabrication. (A) Scanning electron micrograph of the top side and (B) transmission electron micrograph of the center cross-section of a ZMW fabricated by the described photolithography protocol. The desired ZMW shape exhibits vertical bottom walls and a slightly tapered opening toward the top midway up the cladding. Scale bars = 100 nm. (C) Distribution of ZMW floor diameters over an array of 3000 ZMWs with nominal diameter of 100 nm. (mean = 98.4 nm; standard deviation = 5.8 nm).

different materials. It combines selective passivation of the cladding surfaces using polyphosphonate chemistries (Korlach *et al.*, 2008b) with selective functionalization of the ZMW substrate using biotin polyethylene glycol (PEG) silane (Eid *et al.*, 2009). It results in ZMW arrays that present biotin for specific enzyme immobilization only at the ZMW glass floor, above a layer of PEG to preserve enzyme activities (Fig. 20.3A). Using this procedure, we have achieved high contrast ratios of biotin functionalization (in excess of 100:1, Fig. 20.3B), with undetectable levels of biotin PEG silane on the aluminum surface, as measured by X-ray photoelectron spectroscopy (XPS). The specificity of protein binding was also in excess of 100:1 (Fig. 20.3C).

For polymerase immobilization, depending on the ZMW diameter chosen for sequencing, the concentrations of streptavidin and polymerase are adjusted to yield optimal loading which is governed by Poisson statistics (Korlach *et al.*, 2008b) (Section 3.6).

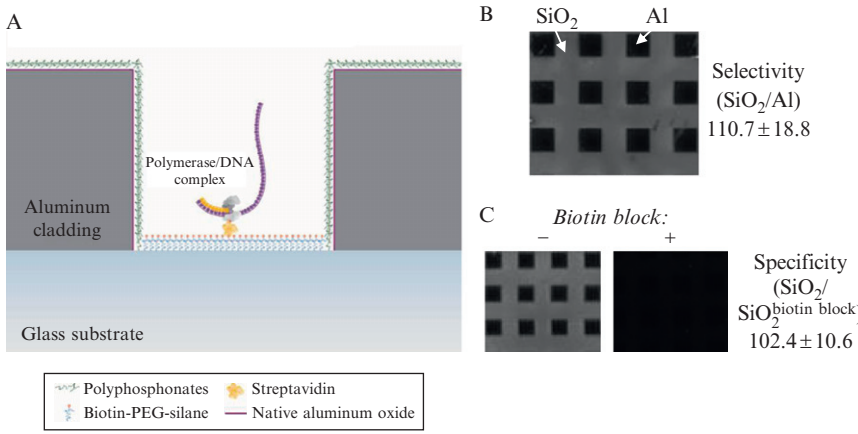

Figure 20.3 Surface derivatization for specific, selective immobilization of DNA polymerase in ZMWs. (A) The cladding surface is passivated with polyphosphonates, the ZMW substrate surface is functionalized with biotin PEG silane to mediate enzyme immobilization targeted to the ZMW floor. (B) Material selectivity of the surface derivatization, measured with 40 nm fluorescent neutravidin beads (Korlach *et al.*, 2008b). The assay uses patterned substrates of 0.5 mm aluminum squares on fused silica. (C) Specificity of neutravidin binding to the glass surface. Specificity was determined by comparing neutravidin binding with a sample for which neutravidin was blocked with excess biotin before the immobilization step. Contrast means and standard deviations are for $n = 4$ chips.

3.3. Phospholinked nucleotides for uninterrupted DNA polymerization

Nucleotides with fluorescent labels attached to the terminal phosphate were first described as efficient substrates for *Escherichia coli* DNA-dependent RNA polymerase (Chatterji and Gopal, 1996; Schlageck *et al.*, 1979; Yarbrough *et al.*, 1979). Utilization of phospholinked dNTPs was subsequently demonstrated in conjunction with DNA polymerases and reverse transcriptases (Kumar *et al.*, 2005; Mulder *et al.*, 2005; Sood *et al.*, 2005), but high concentrations had to be used because incorporation efficiencies were generally much lower than with unmodified dNTPs. It was observed that a linker containing one or more additional phosphate groups extending the triphosphate moiety proved beneficial for improving incorporation efficiencies (Kumar *et al.*, 2005; Sood *et al.*, 2005), presumably due to attenuation of steric hindrance effects of the bulky fluorophore proximal to the active site, and partial restoration of the negative charge lost by the linker conjugation at the γ-phosphate. We have found that for ϕ29 DNA polymerase, a linker that is extending the natural triphosphate by an additional two or three phosphates, followed by a short aminohexyl aliphatic

chain, yields nucleotides that are incorporated as efficiently as unmodified, natural dNTPs with respect to synthesis rates and processivity (Korlach *et al.*, 2008a). Depending on the specific DNA polymerase under study, the exact nature of the linker and number of extra phosphates may vary.

An exemplary synthesis protocol for the phospholinked dNTP Alexa Fluor 488 aminohexyl-dG5P (Korlach *et al.*, 2008a) is given below. Based on carbonyldiimidazole (CDI) activation, it builds the additional diphosphate moiety on an aliphatic linker, allowing for flexibility of dye conjugation as the final step. The aliphatic linker also allows a larger spatial separation between nucleotide and fluorophore. During the third CDI activation for coupling the linker-triphosphate to the nucleotide, $MgCl_2$ is included, which significantly improves the yield (Kadokura *et al.*, 1997).

Fmoc-6-aminohexylphosphate

1. Coevaporate 1 g (2.94 m*M*) of Fmoc-6-aminohexanol 2× with 20 ml anhydrous acetonitrile, then suspend in 10 ml anhydrous triethylphosphate.
2. Add two equivalents (550 μl, 5.88 m*M*) of phosphorus oxychloride to the stirring suspension (Yoshikawa *et al.*, 1967). After 2 h, HPLC shows disappearance of the Fmoc-6-aminohexanol.
3. Quench the reaction by the addition of 100 ml 0.1 *M* triethylamine bicarbonate (TEAB) (pH 6.8) and stir for 30 min. Purify by reverse phase HPLC on a Waters Xterra C18 RP 30 × 100 column using an acetonitrile gradient in 0.1 *M* TEAB.
4. Evaporate the fractions containing product, followed by coevaporation with methanol (2×).
5. Triturate the residue twice with 100 ml diethylether and dry under vacuum to give a white powder. Yield: 1.24 g, 68% as bis-triethylamine salt. Purity (HPLC): 98%.

Fmoc-6-aminohexyldiphosphate

1. Coevaporate 200 mg (320 μ*M*) of Fmoc-6-aminohexylphosphate twice with anhydrous acetonitrile, then suspend in 2 ml anhydrous DMF.
2. Add four equivalents of 1,1′-carbonyldiimidazole (CDI; 207 mg, 1280 μ*M*) and stir at ambient temperature for 4 h (Hoard and Ott, 1965).
3. Add six equivalents of methanol (77 μl, 1920 μ*M*) and stir for 30 min.
4. Add to the reaction 10 equivalents of tributylamine–H_2PO_4 (3200 μ*M*; prepared by mixing equimolar amounts of tributylamine and 85% phosphoric acid, followed by coevaporation three times with anhydrous acetonitrile, and dissolved in 4 ml anhydrous DMF). Stir the reaction mixture for 16 h. HPLC shows 3% Fmoc-aminohexylphosphate remaining.
5. Dilute the reaction mixture to 50 ml with 0.1 *M* TEAB, and purify by RP HPLC on a Waters Xterra C18 RP 30 × 100 column using an acetonitrile gradient in 0.1 *M* TEAB.

6. Evaporate the fractions containing product, followed by coevaporation with methanol (2×).
7. Coevaporate the residue with anhydrous acetonitrile. Yield: 186 mg, 73% as Tris–TEA salt. Purity (HPLC): 96%.

Aminohexyl-dG5P

1. Coevaporate 186 mg (233 μM) Fmoc-6-aminohexyldiphosphate twice with anhydrous acetonitrile, then suspend in 3 ml anhydrous DMF.
2. Add four equivalents of CDI (150 mg, 930 μM) and stir at ambient temperature for 4 h.
3. Add six equivalents of methanol (56 μl, 1400 μM) and stir for 30 min.
4. Coevaporate 1.5 equivalents dGTP (TEA salt, 350 μM) 3× with anhydrous acetonitrile, then suspend in 2 ml anhydrous DMF.
5. Add the Fmoc-aminohexyldiphosphoimidazolate reaction to the dGTP solution, followed by 10 equivalents of anhydrous $MgCl_2$ (333 mg, 3500 μM) (Kadokura *et al.*, 1997). Stir the reaction for 18 h. HPLC shows 28% of the Fmoc-aminohexyldiphosphate converted to Fmoc-aminohexyl-dG5P.
6. Dilute the reaction mixture to 125 ml with 0.1 *M* TEAB, and purify by RP HPLC on a Waters Xterra C18 RP 30 × 100 column using an acetonitrile gradient in 0.1 *M* TEAB.
7. Evaporate the fractions containing product, followed by coevaporation with methanol (2×).
8. Take up the residue in 20 ml 10% TEA/water and stir for 16 h to remove the Fmoc protecting group from the amine on the linker.
9. Evaporate triethylamine, add water to 25 ml, and extract the solution three times with 25 ml diethyl ether.
10. Purify the product from the aqueous layer by anion exchange chromatography on Q sepharose FF using a TEAB gradient from 0.05 to 1 *M*. Yield: 42 μM, 18%. Purity (HPLC): 98%.

Alexa Fluor 488-aminohexyl-dG5P

1. Dissolve 1 μM aminohexyl-dG5P in 200 μl of 50 m*M* $NaHCO_3$, pH 8.7, and add to 1 mg Alexa Fluor 488-TFP ester (Invitrogen). Briefly sonicate the mixture.
2. After 4 h, HPLC showed no active ester remaining (the product is identified by characteristic PDA scan). Purify the compound by IEX on Q sepharose FF with a TEAB gradient from 0.05 to 1 *M*. Purify the product further by RP HPLC on a Waters Xterra RP C18 19 × 100 column using an acetonitrile gradient in 0.1 *M* TEAB.
3. Evaporate the fractions containing pure product, followed by coevaporation with methanol (2×).
4. Dissolve the residue in water and quantitate by UV–Vis spectrophotometry. Yield: 370 n*M*, 37%. Purity (HPLC): 99%.

In contrast to base-linked nucleotide derivatizations, this synthesis scheme proceeds identically for all four nucleobases, and for labeling with different fluorophores using the appropriate dye NHS esters. For phospholinked nucleotides containing a triphosphate linker (Eid *et al.*, 2009) instead of the diphosphate moiety described here, Fmoc-aminohexyl-triphosphate is substituted for Fmoc-aminohexyl-diphosphate in the condensation reaction with nucleoside triphosphate.

Demands on the purity of phospholinked nucleotides for accurate SMRT DNA sequencing are high, as even small traces of unlabeled nucleotides, when incorporated by the polymerase, could lead to missed bases in the sequencing read. Exposure of materials to ambient light should be minimized during the dye conjugation step to avoid fluorophore bleaching. In addition, in contrast to base-linked nucleotides, it is straightforward to subject phospholinked nucleotides to an additional enzymatic purification. For example, one can take advantage of the specificity of alkaline phosphatases which rapidly degrade unmodified dNTPs to the corresponding nucleoside, but are completely inactive on dNTPs that contain moieties coupled to the terminal phosphate (Sood *et al.*, 2005; Yarbrough, 1978). Such post-chemical-synthesis purification can be carried out before using phospholinked nucleotides in SMRT sequencing reactions, or conveniently, the phosphatase can be included in the sequencing reaction.

3.4. DNA polymerase—the sequencing "engine"

Various DNA polymerases can be used in conjunction with SMRT DNA sequencing, and the sequencing performance will depend on their specific properties. We have applied wild-type and mutant DNA polymerases from bacteriophage ϕ29 to our SMRT DNA sequencing method, taking advantage of several favorable characteristics. ϕ29 DNA polymerase is extremely processive (tens of kilobases), relatively fast ($\sim$50–100 bases/s) and highly accurate (error rate of $\sim 10^{-5}$–10^{-6}) (Baner *et al.*, 1998; Blanco *et al.*, 1989; Esteban *et al.*, 1993). It is also very stable, maintaining constant enzymatic activities for up to several days (Dean *et al.*, 2001; Nelson *et al.*, 2002). The use of double-stranded DNA templates is possible by its efficient DNA strand displacement synthesis activity, thus simplifying sample preparation procedures.

The following protocol outlines the expression and purification of $\phi 29^{N62D}$, a mutant with reduced 3′-5′ exonuclease activity while maintaining essentially identical polymerizing properties (Blanco and Salas, 1996; de Vega *et al.*, 1996). N-terminal Histidine (His) and GST tags were cloned into pET41 (Invitrogen, Carlsbad, CA) for ease of purification (Korlach *et al.*, 2008a).

1. Overproduce polymerase in *E. coli* by addition of IPTG (1 m*M*) at midlog phase.

2. Disrupt the cells using lysozyme (chicken egg white, Sigma-Aldrich, St. Louis, MO) in a buffer containing 50 m*M* Tris–HCl, pH 7.5, 7 m*M* 2-mercaptoethanol, and 5% glycerol (buffer B), and additionally containing 0.2 *M* NaCl for 30 min. Sonicate for 2 min with a sonication probe. Degrade DNA with DNase I (bovine pancreas, Sigma-Aldrich) for 30 min at room temperature while shaking. Remove cell debris by centrifugation for 30 min at 15,000×*g*.
3. Adjust the supernatant to 0.5 *M* NaCl for purification on a 1-ml HisTrap FF column (Ni-resin, GE Healthcare, Piscataway, NJ), equilibrated with buffer B containing 0.5 *M* NaCl. Wash His-tagged polymerase retained on the HisTrap column with at least 50 column volumes, first using buffer B containing 1 *M* NaCl, 0.2% Tween-20, and 20 m*M* imidazole, followed by buffer B containing 0.5 *M* NaCl and 50 m*M* imidazole.
4. Elute the polymerase with buffer B containing 300 m*M* imidazole. Pool this "HisTrap" fraction and adjust it to 0.2 *M* NaCl using buffer C (50 m*M* Tris–HCl, pH 7.5, 1 m*M* EDTA, 7 m*M* 2-mercaptoethanol, 5% (v/v) glycerol).
5. Load the sample onto a Heparin-Sepharose CL-6B column (10 ml, GE Healthcare), equilibrated in buffer C. Elute polymerase with buffer C and a gradient of 0–1 *M* NaCl. Pool these "Heparin" fractions and concentrate using Centricon YM-50 (Millipore, Billerica, MA).
6. Adjust to the final storage buffer (50 m*M* Tris–HCl, pH 7.5, 0.2 *M* NaCl, 1 m*M* EDTA, 7 m*M* 2-mercaptoethanol, 50% (v/v) % glycerol).
7. Analyze protein fractions by sodium dodecyl sulfate—polyacrylamide gel electrophoresis (SDS–PAGE, 10% or 15% polyacrylamide). The purified polymerase is at least 97% homogenous. Determine protein concentrations both by measuring the absorbance at 280 nm, and by the Bradford method using known amount of BSA (Bio-Rad).

Various strategies exist to mediate specific enzyme immobilization in ZMWs. They will depend on the particular coupling chemistry chosen. For a surface derivatization providing selectively biotinylated ZMW floors (described in Section 3.2), they include nonspecific biotinylation of the enzyme (Hermanson, 1996), introduction of specific biotin tags (Beckett *et al.*, 1999), or fusion proteins (Nilsson *et al.*, 1997).

3.5. Instrument for highly parallel monitoring of sequencing reactions

As for all single-molecule fluorescence recording systems, optimization of light collection efficiencies is paramount for maximal signal-to-noise detection. For SMRT DNA sequencing, additional demands exist for simultaneously illuminating and monitoring many ZMWs containing immobilized DNA polymerase molecules. We have described the development of a

highly parallel optical system that is capable of continuously analyzing thousands of concurrent sample locations (Lundquist *et al.*, 2008) (Fig. 20.4).

In this system, wavelength-specific holographic phase masks (Kress and Meyrueis, 2000) act as illumination multiplexers by dividing the laser beams into several thousand subbeams. Relay lens assemblies convert these beams into corresponding arrays of spots focused at a plane conjugate to the front focus of a microscope objective. After combining multiple wavelengths paths, all of the illumination light is transmitted through a common dichroic filter and brought to an array of diffraction-limited focal spots in the sample plane where it excites fluorescence in each ZMW observation volume.

Multiple laser sources, coupled with demands for detecting several different fluorophores with single-molecule sensitivity, put stringent demands on the dichroic filter. We have found that better performance

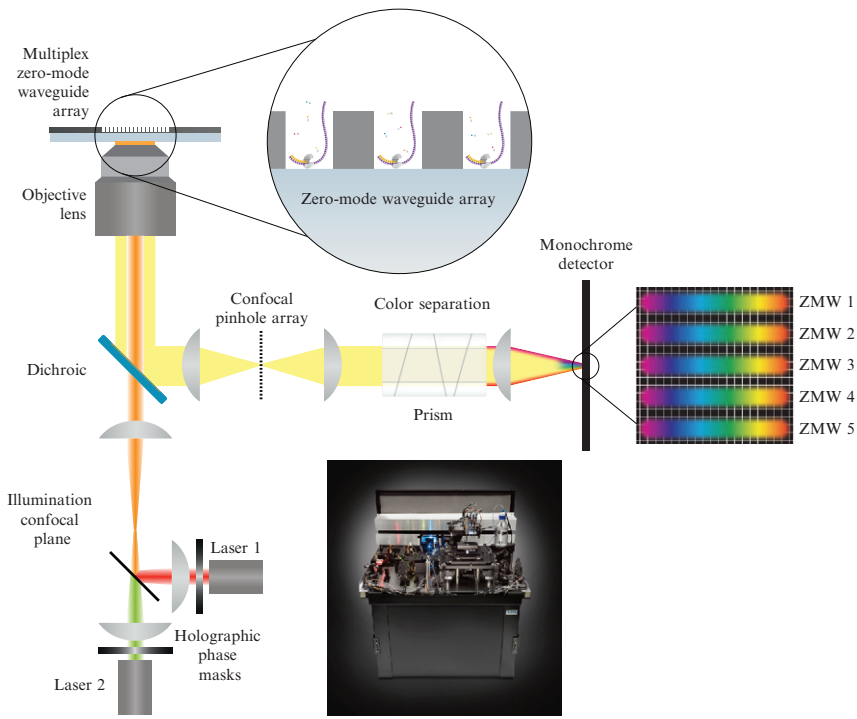

Figure 20.4 Schematics of an optical system for SMRT DNA Sequencing. This instrument provides simultaneous illumination of 3000 ZMWs with two different lasers, and wavelength-specific real-time detection of fluorescence from phospholinked nucleotides processed by DNA polymerase immobilized in the ZMWs. A photograph of an instrument (top cover removed) is shown on the bottom.

with respect to narrow excitation bandwidths and high fluorescence throughput can be achieved by reversing the conventional positions of the illumination and collection paths.

Upon reflecting off the dichroic filter, fluorescence light is first imaged onto a congruent array of confocal pinholes to reject stray light not originating from ZMW locations. It is subsequently reimaged through a prism dispersive element onto a monochrome, electron-multiplication charge-coupled device (EM-CCD) array. The compound prism serves to linearly disperse the wavelength content of the emitted light, transforming the image from each confocal volume element from a diffraction-limited spot into a "rainbow" pattern. Emitted fluorescence from different phospholinked nucleotides falls onto different spatial locations on the detector, thereby enabling identification of the type of nucleotide incorporated by the polymerase at any given time. A single camera thus collects both spatial and spectral information for the entire ZMW array. The prism assembly we have chosen allows for high-transmission, continuous color separation. The resulting opportunity for oversampling the fluorescence spectra in wavelength improves the accuracy of classification in cases of overlapping dye emission spectra. For example, in the implementation described in Lundquist *et al.*, (2008), a three-wedge compound prism was optimized to provide linear angular dispersion of 1.25 mrad between the wavelengths of 490 and 730 nm, with a zero-deviation angle at 550 nm. With this choice, the system can be applied to a variety of fluorophore combinations. If resolution of highly overlapping emission spectra were a limitation, dispersion could be increased in the critical regions to improve spectral performance.

3.6. DNA sequencing assay example

A typical protocol for performing a SMRT DNA sequencing reaction is described below. An oxygen scavenging system, consisting of protocatechuate dioxygenase (PCD) and its substrate protocatechuic acid (PCA), is used to remove oxygen and thereby suppress photophysical effects that might be deleterious to the polymerase (Aitken *et al.*, 2008; Eid *et al.*, 2009). A nitrobenzyl-based triplet state quencher, nitrobenzoic acid (NBA), is included to shorten the time fluorophores can reside in the dark triplet state (Dave *et al.*, 2009).

1. Incubate DNA polymerase carrying an N-terminal biotin-tag (Beckett *et al.*, 1999) with 1.5× molar excess of primed DNA template at 4 °C for 10 min in a buffer containing 50 m*M* MOPS, pH 7.5, 75 m*M* potassium acetate, 5 m*M* dithiothreitol and 0.05% (v/v) Tween-20.
2. Simultaneously, incubate streptavidin (Invitrogen) at a twofold stoichiometric excess over polymerase in the same buffer at 22 °C on the ZMW array. Wash the array five times with buffer.

3. Immobilize polymerase/template complexes onto the arrays at 4 °C for 15 min. Remove unbound complexes by washing five times with reaction buffer (50 m*M* ACES, pH 7.1, 75 m*M* potassium acetate, and 5 m*M* dithiothreitol).
4. Add the oxygen scavenging system (1× PCD, 4 m*M* PCA), triplet state quencher (6 m*M* NBA (all Sigma-Aldrich, St. Louis, MO), and all phospholinked dNTPs (250 n*M* final concentration each), except the one corresponding to the first base to be incorporated into the DNA template.
5. Initiate the sequencing reactions by simultaneous addition of the first phospholinked dNTP to be incorporated (250 n*M*) and manganese acetate (0.5 m*M* final concentration) (Korlach *et al.*, 2008a; Kumar *et al.*, 2005).

3.7. Data analysis

Fluorescence pulse calling is performed by a threshold algorithm on the dye-weighted intensities using fluorescence emission calibration spectra for each of the phospholinked dNTPs (Horne, 1986). For base identification, light collected by the detector is summed over the duration of the pulse, allowing each fluorescent phospholinked nucleotide to be spectrally evaluated and classified. Automated classification is performed by least-squares fitting to the four known dye reference spectra. The spectrum that yields the minimum chi-squared difference when compared to the pulse spectrum identifies the pulse as a particular phospholinked dNTP. The degree of spectral cross talk between different fluorophore types depends on the specific dyes chosen and their intrinsic brightness. A typical set of dyes we have used in conjunction with SMRT DNA sequencing, with spectral separation between the emission maxima of the pairs of dyes excited with the same laser of 31 and 23 nm, respectively, showed misidentification rates of less than 1% using this method (Eid *et al.*, 2009).

A section of a typical sequencing read after such automated base calling is shown in Fig. 20.5A. DNA polymerase activity is marked by a train of pulses corresponding to phospholinked nucleotide incorporations. Incorporation signals from the four phospholinked nucleotides show fluorescence intensity "level setting" characteristics which are due to

(i) the stationary location of the enzyme's active site with respect to the illumination profile, and
(ii) different excitation efficiencies and intrinsic fluorophore brightness of the four fluorescent dyes.

High detector frame rates (e.g., 100 Hz; Eid *et al.*, 2009) ensure that enzymatic turnovers are oversampled in time, allowing precise measurements

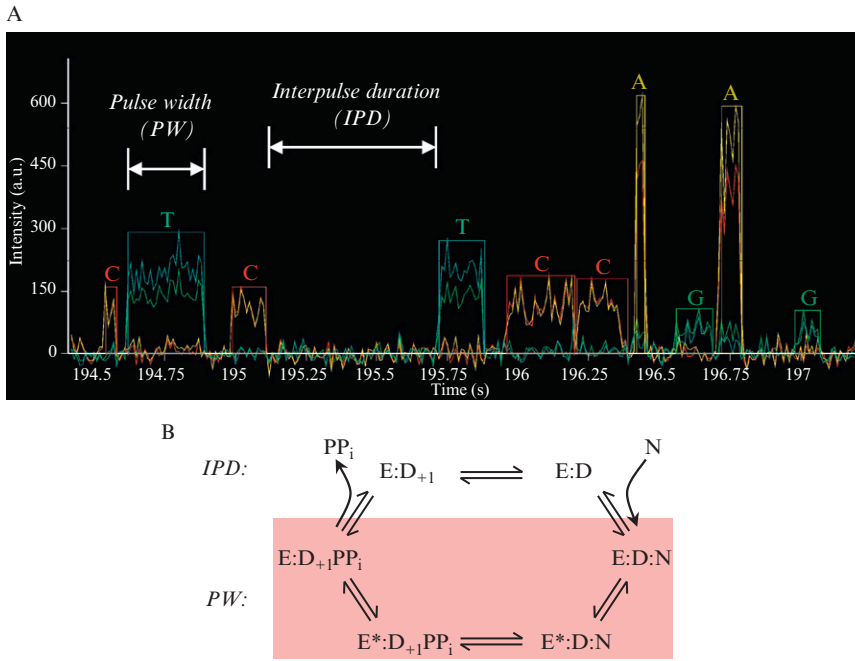

Figure 20.5 SMRT DNA sequencing example. (A) Four-color time trace of dye-weighted fluorescence intensities with automated base annotations, along with definitions of pulse width (PW) and interpulse duration (IPD). While significant spectral overlap can be present from two dyes excited by one laser, its magnitude is known from the dye reference spectra and thus it does not affect misidentification rates. (B) Generalized enzymatic reaction cycle of DNA polymerization, with the red rectangle differentiating the steps which constitute the "bright" (PW) state. E: polymerase enzyme; D: DNA template; N: nucleotide.

of "on" and "off" times. The pulses exhibit stochastic intensity fluctuations because of counting statistics and dye photophysics. Hallmarks of single-molecule fluorescent events are characteristic: single-frame rise and fall times at the start and end of the pulse, respectively ($\ll 10$ ms), which facilitate pulse detection and base calling even when pulses are close together.

Single-molecule events corresponding to phospholinked dNTP incorporations manifest as fluorescent pulses whose variable duration and spacing directly reflect the underlying enzyme kinetics. In the generalized enzymatic DNA polymerization cycle (Fig. 20.5B), kinetic steps starting with phospholinked nucleotide binding, then proceeding through the transition from the "open" to the "closed" polymerase conformation, catalysis, reverse conformational transition, and up to the moment of pyrophosphate-linker-fluorophore release, have the nucleotide bound in the active site and thereby

make up the duration of the fluorescence pulse (defined hereafter as pulse width (PW)). The time elapsing after release of the reaction product, translocation of the polymerase along the template by one base, and waiting time for the binding of the next incoming phospholinked nucleotide constitute the duration between successive pulses (interpulse duration (IPD)).

4. Single-Molecule DNA Polymerase Dynamics

While the base sequence of the synthesized DNA strand constitutes the main output of the SMRT sequencing method, the real-time aspect of this approach generates unprecedented information about DNA polymerase kinetics. Because the system reports the kinetics of every base incorporation through PW and IPD, it can be used to investigate the dynamics of DNA polymerization with base-pair resolution, and to provide the distribution of kinetic parameters over many different sequence contexts in a single 5-min experiment. The method thereby allows direct assessments of static variation (differences in enzymatic activity between different molecules) and dynamic variation (fluctuation of catalytic rate constants over time for a single enzyme molecule).

4.1. Determination of single-molecule kinetic parameters

From the multitude of nucleotide incorporations recorded for single DNA polymerases during a SMRT sequencing read, it is possible to determine effective kinetic parameters for each enzyme molecule. For certain polymerases and reaction conditions, a simplified model can successfully be used for which single steps limit the rates of transitions between the "bright" and "dark" states, as defined above (Fig. 20.5B). In this case, the kinetic cycle of DNA synthesis reduces to

$$\mathrm{N} + \mathrm{E:D} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{E:D:N} \xrightarrow{k_2} \mathrm{E:D}_{+1}$$

with the ternary complex of enzyme, DNA, and nucleotide (E:D:N) now denoting all states for which the phospholinked nucleotide is bound in the active site. Here, the average fluorescence PW is determined by the summation of kinetic steps exiting the "bright" state

$$\mathrm{PW} = \frac{1}{k_{\mathrm{PW}}} = \frac{1}{(k_2 + k_{-1})} \tag{20.1}$$

The IPD is governed by the combination of DNA polymerase translocation along the template and time to binding of the next nucleotide. For enzymes with fast translocation kinetics relative to nucleotide binding (e.g., below nucleotide substrate saturation concentrations), the latter becomes rate-limiting, and the average IPD is then given by

$$\mathrm{IPD} = \frac{1}{k_{\mathrm{IPD}}} = \frac{1}{k_1[N]} \tag{20.2}$$

The two expressions are combined to yield a measurement of the K_m for a single polymerase molecule from measured average PWs and IPDs:

$$K_m = \frac{(k_2 + k_{-1})}{k_1} = \frac{\mathrm{IPD}}{\mathrm{PW}}[N] \tag{20.3}$$

Base-specific K_m values determined in this way for a 10-min, 2.4-kb sequencing read are depicted in Fig. 20.6. The single-exponential fits to the PW and IPD histograms indicate that this simplified, single rate-limiting kinetic model is a good approximation for the DNA polymerase employed in this experiment. K_m values for this molecule are in good agreement with bulk data measuring the transient kinetics of single nucleotide turnovers using a stopped-flow instrument (KinTek), except for dTTP which has a lower K_m for the single polymerase by a factor of ~ 2.

4.2. DNA polymerase pausing

The above-mentioned analysis does not account for transient changes in DNA polymerization dynamics *during* the course of DNA synthesis for a single enzyme. While the validity of the Michaelis–Menten equation has been demonstrated theoretically and experimentally in the presence of such dynamic variation (English *et al.*, 2006; Kou *et al.*, 2005; Min *et al.*, 2006; Velonia *et al.*, 2005), the extracted parameters do not inform about it, so a more detailed analysis is needed. DNA polymerases present a more complex case compared to other, single-substrate turnover enzymes, as there are four different nucleotides competing for one active site, several different enzymatic activities are possible in one protein, and the sequence of the DNA template introduces a large number of different sequence contexts. A detailed description of these phenomena is beyond the scope of this chapter, and only a few examples illustrating the richness of these single-molecule data are given here.

We have previously described polymerase pausing caused by hairpin formation in single-stranded DNA templates (Eid *et al.*, 2009). Here, we extend our analysis to double-stranded DNA templates several hundred

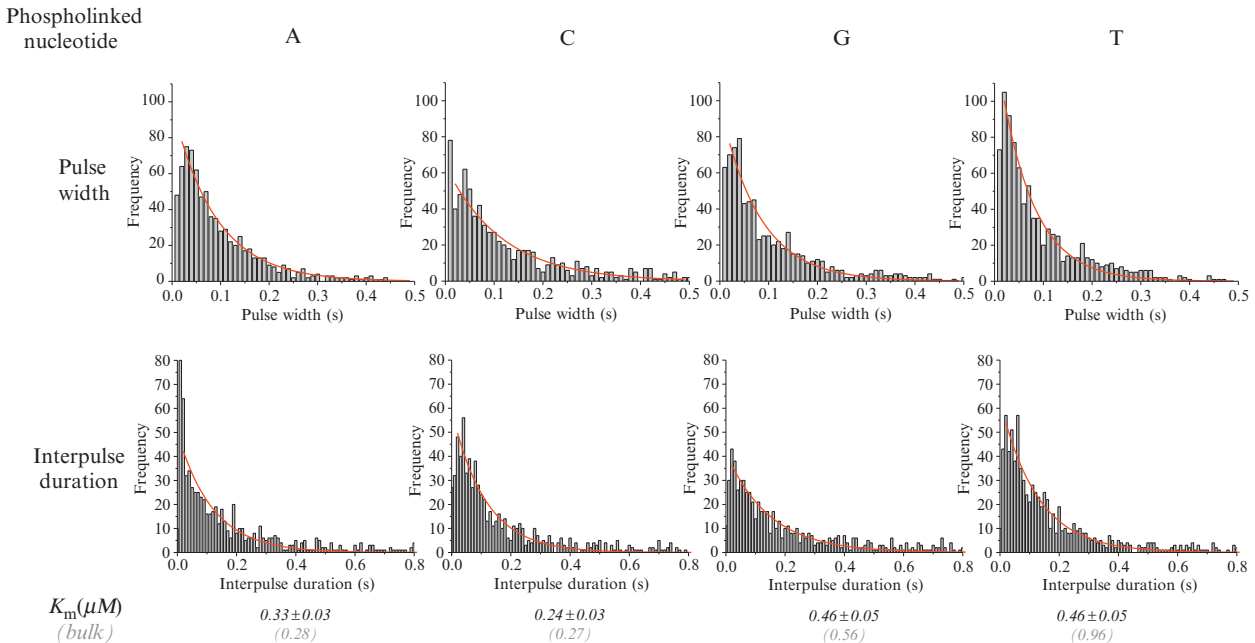

Figure 20.6 Determination of effective K_m values from a SMRT sequencing read. Distributions from a 10-min, 2.4-kb read for pulse widths of, and interpulse duration before each of the four phospholinked nucleotides are fit to single exponential decay functions used to derive K_m values (Eq. (20.3)). Values are compared to bulk measurements of single-turnover kinetics (gray).

bases in length, in conjunction with DNA strand displacement synthesis carried out by the polymerase. While double-stranded DNA is characterized by dramatically less secondary structure in comparison with single-stranded DNA, there still remains a degree of higher order structure due to different sequence contexts (Gimenes *et al.*, 2008; Hagerman, 1990). In addition, the dynamics of strand displacement activity of the polymerase, separating the nontemplate strand from the template strand before its entry into the polymerase active site, is also likely influencing DNA polymerization rates (Kamtekar *et al.*, 2004; Rodriguez *et al.*, 2005).

Single-molecule sequencing trajectories of template position over time allow a detailed view into polymerase-mediated DNA synthesis dynamics (Fig. 20.7). Variations in polymerization rates are apparent with respect to heterogeneity among different molecules for a given template position, allowing the creation of instantaneous rate distributions for every template position when analyzed over all molecules. Template sequence context effects over many time scales are apparent. For the specific example shown in Fig. 20.7A, the DNA synthesis rate roughly doubles after incorporation of $\sim$250 nucleotides. Analyzed over shorter time scales, a range of interpulse distances is observed, including occurrences of cessation of DNA synthesis activity on a time scale of many seconds, followed by resumption of DNA synthesis. Such polymerase pausing is specific to certain DNA template locations where it is common to several molecules (Fig. 20.7A, arrows).

To rule out that such variations are caused by systematic effects of the instrumentation or reaction conditions, and instead represent dynamic changes of DNA polymerization caused by different DNA template sequence contexts, *circular* DNA templates can be employed (Eid *et al.*, 2009). In this configuration, polymerases capable of strand displacement synthesis, such as ϕ29 DNA polymerase (Blanco *et al.*, 1989), will encounter the same DNA sequence on a template molecule multiple times. DNA synthesis trajectories encompassing multiple rounds of circular template sequencing (Fig. 20.7B) show that polymerases repeat the dynamic signatures upon consecutive encounters of the same DNA template sequence context. For example, the transition to the faster overall rate at around 250 nucleotides into the template (described above) can be discerned for each round of synthesis. Similarly, polymerase pausing occurs at the same DNA template site during each round of synthesis, albeit with variable IPDs (Fig. 20.7B, lower right panel, arrows). To within the resolution of the data in this experiment, the correlation between IPDs of the same sequence context in different molecules was the same as between different laps around the circular template carried out by the same molecule. IPD profiles from several hundred polymerases for multiple laps illustrate these phenomena further (Fig. 20.7C), showing the slower rate for the first half of the template due to overall increased IPDs, and highlighting template locations of increased pausing frequencies.

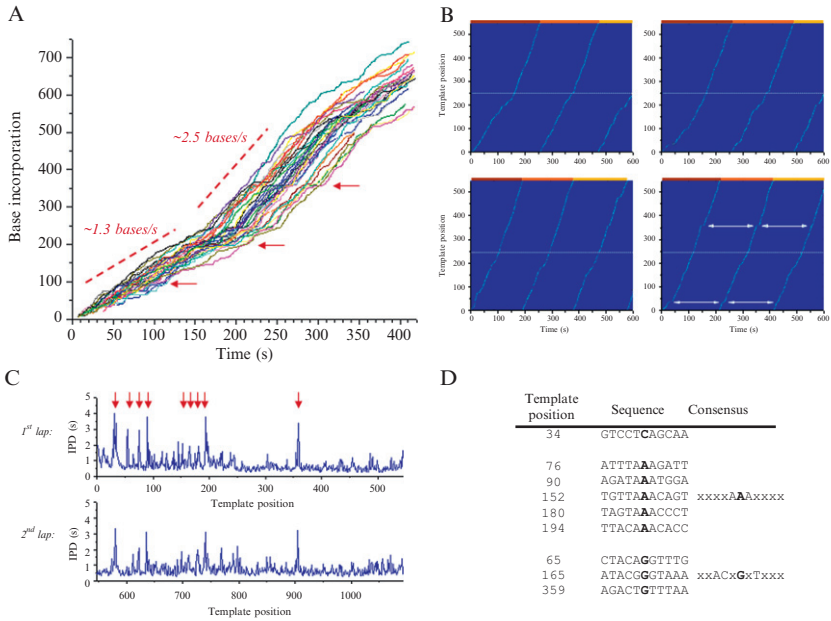

Figure 20.7 SMRT DNA synthesis dynamics. (A) DNA synthesis trajectories of template position over time for 33 polymerase molecules. The two dashed lines illustrate a transition to higher polymerization rates occurring at ~250 bases into the DNA template. Arrows indicate common sites in the DNA template at which pausing occurs for several enzyme molecules. (B) Polymerization trajectories for four DNA polymerase molecules on circular DNA templates, showing multiple laps of continuous, processive DNA synthesis. The white dotted line indicates the same template position as in (A) where the speed transition occurs. Arrows in the lower right panel show an example of template locations at which the polymerase molecule pauses during each round of synthesis. (C) 90th percentile of the IPD versus template position for several hundred polymerases sequencing two successive laps around a circular DNA template. The arrows indicate the pause sites for which the sequences and consensus contexts (D) are given.

Analysis of DNA sequences for these pauses indicate common, relatively short consensus sequence contexts (Fig. 20.7D) which we have confirmed using templates from different sources (data not shown). While the mechanism and biological significance of these pausing signatures remains to be elucidated, the ''AAA'' sequence is reminiscent of the initiation mechanism by ϕ29 DNA polymerase (characterized by multiple dATP incorporations and sliding-back movements; Mendez *et al.*, 1992). We speculate that perhaps the polymerase has an increased propensity for slippage in this template sequence context, creating the pause signature before DNA synthesis resumes.

5. Conclusions

SMRT DNA sequencing harnesses the intrinsic power of DNA polymerases, allowing their speed, processivity, efficiency, and fidelity to be exploited directly. Rapid intrinsic DNA synthesis rates translate to short sequencing run cycle times. Long continuous sequence reads preserve the molecular integrity of the DNA template, simplifying the downstream bioinformatics for genome assembly and analysis in the context of structural variations and allelic polymorphism linkages.

Many sequencing-by-synthesis techniques employ DNA polymerase as a bulk reagent consumable, synchronizing its activity with various termination approaches (reviewed in Mardis, 2008). Such gating allows for an increase in multiplex and overall sequence throughput, but comes at a cost of long single-base cycle times and relatively short read lengths due to incomplete cycle efficiencies. These methods utilize uniform protocols for each incorporation cycle and are therefore insensitive to sequence context effects on polymerization efficiencies, leading to variable systematic errors (Kong, 2009) and even inaccessible genomic regions. In contrast, sequencing by observing uninterrupted DNA polymerization allows the enzyme to spend variable amounts of

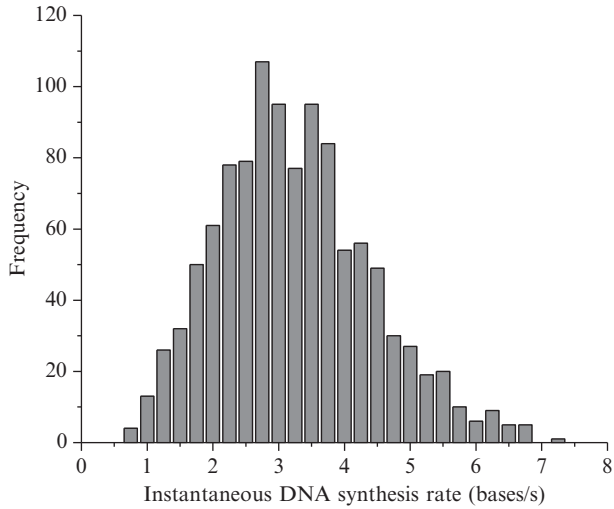

Figure 20.8 DNA polymerization rate variations over different sequence contexts. In this example, SMRT sequencing reads were recorded from 128 polymerase molecules. Instantaneous DNA synthesis rates for each position of the 1.1 kb DNA template were extracted for each molecule. Mean values from the rate distributions at each template position are used for the histogram. The coefficient of variation of DNA synthesis rate for this template is 70%.

time synthesizing different sequence contexts. Even when analyzed over many polymerase molecules, significant variation of DNA synthesis rates exists for different DNA template positions (Fig. 20.8). This dispersion of DNA synthesis rates as a function of sequence context suggests that in an ensemble sequencing-by-synthesis system, the ideal protocol for incorporation should vary by that same amount. This requirement is moot in a single-molecule, real-time method, as every incorporation is allowed precisely the correct amount of time, thus providing an intrinsic adjustment and optimal sequencing performance in every sequence context.

Heterogeneities between enzymes (static variation) or in the catalytic rates for a single enzyme molecule (dynamic variation) have been observed for many enzymes (reviewed in Blank *et al.*, 2009), and DNA polymerases are no exception. The information can be used advantageously to improve the quality of sequencing, and at the same time provide insights into the dynamics of DNA polymerization. Because polymerase kinetics is sensitive to biological perturbation, this information can be further developed for investigating DNA binding proteins, DNA polymerase inhibitors, and effects of base methylation.

ACKNOWLEDGMENTS

We are indebted to the entire staff at Pacific Biosciences for their dedicated work that brings this technology to fruition. We also thank J. Puglisi, M. Hunkapiller, R. Kornberg, K. Johnson, D. Haussler, W. Webb, and H. Craighead for many helpful discussions. Aspects of this research were supported by National Human Genome Research Institute grant R01HG003710.

REFERENCES

Aitken, C. E., Marshall, R. A., and Puglisi, J. D. (2008). An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **94,** 1826–1835.

Baner, J., Nilsson, M., Mendel-Hartvig, M., and Landegren, U. (1998). Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res.* **26,** 5073–5078.

Beckett, D., Kovaleva, E., and Schatz, P. J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8,** 921–929.

Blanco, L., Bernad, A., Lazaro, J. M., Martin, G., Garmendia, C., and Salas, M. (1989). Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J. Biol. Chem.* **264,** 8935–8940.

Blanco, L., and Salas, M. (1996). Relating structure to function in phi29 DNA polymerase. *J. Biol. Chem.* **271,** 8509–8512.

Blank, K., De Cremer, G., and Hofkens, J. (2009). Fluorescence-based analysis of enzymes at the single-molecule level. *Biotechnol. J.* **4,** 465–479.

Chatterji, D., and Gopal, V. (1996). Fluorescence spectroscopy analysis of active and regulatory sites of RNA polymerase. *Methods Enzymol.* **274,** 456–478.

Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G., Robertson, A. J., Perkins, A. C., *et al.* (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **5,** 613–619.

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., and Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452,** 215–219.

Dave, R., Terry, D. S., Munro, J. B., and Blanchard, S. C. (2009). Mitigating unwanted photophysical processes for improved single-molecule fluorescence imaging. *Biophys. J.* **96,** 2371–2381.

de Vega, M., Lazaro, J. M., Salas, M., and Blanco, L. (1996). Primer-terminus stabilization at the 3′-5′ exonuclease active site of phi29 DNA polymerase. Involvement of two amino acid residues highly conserved in proofreading DNA polymerases. *EMBO J.* **15,** 1182–1192.

Dean, F. B., Nelson, J. R., Giesler, T. L., and Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11,** 1095–1099.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., *et al.* (2009). Real-time DNA sequencing from single polymerase molecules. *Science* **323,** 133–138.

English, B. P., Min, W., van Oijen, A. M., Lee, K. T., Luo, G., Sun, H., Cherayil, B. J., Kou, S. C., and Xie, X. S. (2006). Ever-fluctuating single enzyme molecules: Michaelis–Menten equation revisited. *Nat. Chem. Biol.* **2,** 87–94.

Esteban, J. A., Salas, M., and Blanco, L. (1993). Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J. Biol. Chem.* **268,** 2719–2726.

Foquet, M., Samiee, K. T., Kong, X. X., Chauduri, B. P., Lundquist, P. M., Turner, S. W., Freudenthal, J., and Roitman, D. B. (2008). Improved fabrication of zero-mode waveguides for single-molecule detection. *J. Appl. Phys.* **103,** 034301.

Fullwood, M. J., Wei, C. L., Liu, E. T., and Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.* **19,** 521–532.

Gimenes, F., Takeda, K. I., Fiorini, A., Gouveia, F. S., and Fernandez, M. A. (2008). Intrinsically bent DNA in replication origins and gene promoters. *Genet. Mol. Res.* **7,** 549–558.

Hagerman, P. J. (1990). Sequence-directed curvature of DNA. *Annu. Rev. Biochem.* **59,** 755–781.

Hermanson, G. T. (1996). Bioconjugate Techniques. Academic Press, San Diego.

Hoard, D. E., and Ott, D. G. (1965). Conversion of mono- and oligodeoxyribonucleotides to 5′-triphosphates. *J. Am. Chem. Soc.* **87,** 1785–1788.

Horne, K. (1986). An optimal extraction algorithm for CCD spectroscopy. *Publ. Astron. Soc. Pac.* **98,** 609–617.

Kadokura, M., Wada, T., Urashima, C., and Sekine, M. (1997). Efficient synthesis of gamma-methyl-capped guanosine 5′-triphosphate as a 5′-terminal unique structure of U6 RNA via a new triphosphate bond formation involving activation of methyl phosphorimidazolidate using ZnCl2 as a catalyst in DMF under anhydrous conditions. *Tetrahedron Lett.* **38,** 8359–8362.

Kamtekar, S., Berman, A. J., Wang, J., Lazaro, J. M., de Vega, M., Blanco, L., Salas, M., and Steitz, T. A. (2004). Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage phi29. *Mol. Cell* **16,** 609–618.

Kong, Y. (2009). Statistical distributions of sequencing by synthesis with probabilistic nucleotide incorporation. *J. Comput. Biol.* **16,** 817–827.

Korlach, J., Bibillo, A., Wegener, J., Peluso, P., Pham, T. T., Park, I., Clark, S., Otto, G. A., and Turner, S. W. (2008a). Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides Nucleotides Nucleic Acids* **27,** 1072–1083.

Korlach, J., Marks, P. J., Cicero, R. L., Gray, J. J., Murphy, D. L., Roitman, D. B., Pham, T. T., Otto, G. A., Foquet, M., and Turner, S. W. (2008b). Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proc. Natl. Acad. Sci. USA* **105,** 1176–1181.

Kou, S. C., Cherayil, B. J., Min, W., English, B. P., and Xie, X. S. (2005). Single-molecule Michaelis–Menten equations. *J. Phys. Chem. B* **109,** 19068–19081.

Kress, B., and Meyrueis, P. (2000). Digital diffractive optics: An introduction to planar diffractive optics and related technology. John Wiley, Chichester; New York.

Kumar, S., Sood, A., Wegener, J., Finn, P. J., Nampalli, S., Nelson, J. R., Sekher, A., Mitsis, P., Macklin, J., and Fuller, C. W. (2005). Terminal phosphate labeled nucleotides: Synthesis, applications, and linker effect on incorporation by DNA polymerases. *Nucleosides Nucleotides Nucleic Acids* **24,** 401–408.

Levene, M. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G., and Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299,** 682–686.

Liu, Y., and Blair, S. (2003). Fluorescence enhancement from an array of subwavelength metal apertures. *Opt. Lett.* **28,** 507–509.

Lundquist, P. M., Zhong, C. F., Zhao, P. Q., Tomaney, A. B., Peluso, P. S., Dixon, J., Bettman, B., Lacroix, Y., Kwo, D. P., McCullough, E., Maxham, M., Hester, K., *et al.* (2008). Parallel confocal detection of single molecules in real time. *Opt. Lett.* **33,** 1026–1028.

Maher, C. A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N., and Chinnaiyan, A. M. (2009). Transcriptome sequencing to detect gene fusions in cancer. *Nature* **458,** 97–101.

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Ann. Rev. Genomics Hum. Genet.* **9,** 387–402.

Mendez, J., Blanco, L., Esteban, J. A., Bernad, A., and Salas, M. (1992). Initiation of phi 29 DNA replication occurs at the second 3′ nucleotide of the linear template: A sliding-back mechanism for protein-primed DNA replication. *Proc. Natl. Acad. Sci. USA* **89,** 9579–9583.

Min, W., Gopich, I. V., English, B. P., Kou, S. C., Xie, X. S., and Szabo, A. (2006). When does the Michaelis–Menten equation hold for fluctuating enzymes? *J. Phys. Chem. B* **110,** 20093–20097.

Miyake, T., Tanii, T., Sonobe, H., Akahori, R., Shimamoto, N., Ueno, T., Funatsu, T., and Ohdomari, I. (2008). Real-time imaging of single-molecule fluorescence with a zero-mode waveguide for the analysis of protein–protein interaction. *Anal. Chem.* **80,** 6018–6022.

Mulder, B. A., Anaya, S., Yu, P., Lee, K. W., Nguyen, A., Murphy, J., Willson, R., Briggs, J. M., Gao, X., and Hardin, S. H. (2005). Nucleotide modification at the gamma-phosphate leads to the improved fidelity of HIV-1 reverse transcriptase. *Nucleic Acids Res.* **33,** 4865–4873.

Nelson, J. R., Cai, Y. C., Giesler, T. L., Farchaus, J. W., Sundaram, S. T., Ortiz-Rivera, M., Hosta, L. P., Hewitt, P. L., Mamone, J. A., Palaniappan, C., and Fuller, C. W. (2002). TempliPhi, phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. *Biotechniques* (Suppl.), 44–47.

Nilsson, J., Stahl, S., Lundeberg, J., Uhlen, M., and Nygren, P. A. (1997). Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Expr. Purif.* **11,** 1–16.

Rigneault, H., Capoulade, J., Dintinger, J., Wenger, J., Bonod, N., Popov, E., Ebbesen, T. W., and Lenne, P.-F. (2005). Enhancement of single-molecule fluorescence detection in subwavelength apertures. *Phys. Rev. Lett.* **95,** 117401.

Rodriguez, I., Lazaro, J. M., Blanco, L., Kamtekar, S., Berman, A. J., Wang, J., Steitz, T. A., Salas, M., and de Vega, M. (2005). A specific subdomain in phi29 DNA polymerase confers both processivity and strand-displacement capacity. *Proc. Natl. Acad. Sci. USA* **102,** 6407–6412.

Schlageck, J. G., Baughman, M., and Yarbrough, L. R. (1979). Spectroscopic techniques for study of phosphodiester bond formation by *Escherichia coli* RNA polymerase. *J Biol. Chem.* **254,** 12074–12077.

Sood, A., Kumar, S., Nampalli, S., Nelson, J. R., Macklin, J., and Fuller, C. W. (2005). Terminal phosphate-labeled nucleotides with improved substrate properties for homogeneous nucleic acid assays. *J. Am. Chem. Soc.* **127,** 2394–2395.

Tabor, S., Huber, H. E., and Richardson, C. C. (1987). *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *J. Biol. Chem.* **262,** 16212–16223.

Velonia, K., Flomenbom, O., Loos, D., Masuo, S., Cotlet, M., Engelborghs, Y., Hofkens, J., Rowan, A. E., Klafter, J., Nolte, R. J., and de Schryver, F. C. (2005). Single-enzyme kinetics of CALB-catalyzed hydrolysis. *Angew. Chem. Int. Ed. Engl.* **44,** 560–564.

Yarbrough, L. R. (1978). Synthesis and properties of a new fluorescent analog of ATP: Adenosine-5′-triphosphoro-gamma-1-(5-sulfonic acid) napthylamidate. *Biochem. Biophys. Res. Commun.* **81,** 35–41.

Yarbrough, L. R., Schlageck, J. G., and Baughman, M. (1979). Synthesis and properties of fluorescent nucleotide substrates for DNA-dependent RNA polymerases. *J. Biol. Chem.* **254,** 12069–12073.

Yassour, M., Kaplan, T., Fraser, H. B., Levin, J. Z., Pfiffner, J., Adiconis, X., Schroth, G., Luo, S., Khrebtukova, I., Gnirke, A., Nusbaum, C., Thompson, D. A., *et al.* (2009). Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. *Proc. Natl. Acad. Sci. USA* **106,** 3264–3269.

Yoshikawa, M., Kato, T., and Takenishi, T. (1967). A novel method for phosphorylation of nucleosides to 5′-nucleotides. *Tetrahedron Lett.* **50,** 5065–5068.