Genomics

DOI: 10.1002/anie.200900722

Sequencing Single DNA Molecules in Real Time

*Jenny Ibach and Susanne Brakmann**

DNA polymerase · DNA sequencing · fluorescence · nucleotides · single-molecule detection

The market for fast and inexpensive DNA sequencing technology has never been larger than today. It comprises research interests like comparative genomics, evolution, and epidemiology studies as well as forensic examinations and the large sector of medical diagnostics. Above all, identification of human DNA sequence variation promises a key to understanding health and disease. The strong interest in these biomedical problems has triggered the development of diverse techniques[1a] for single-nucleotide polymorphism (SNP) genotyping or DNA sequencing in real-time and with the potential to be massively scaled up. An excellent example is pyrosequencing,[1b] which is limited in its ability to sequence whole genomes but is ideally suited for sequencing small regions with read lengths of less than 250 bases to perform SNP genotyping and to enable the identification of multiple SNPs in a small region.

The development of next-generation sequencing technologies is intended to overcome the immense expenses of material, labor, and time associated with amplification, cloning, and purification of DNA material for current sequencing technologies. The ultimate goal is to bring down the cost of whole-genome sequencing to $1000.[2] Strategies currently being pursued include sequencing by hybridization (SBH), nanopore sequencing, sequencing by synthesis (SBS), and the reverse process, sequencing by degradation.[3] Among the diverse approaches, those strategies that involve the direct observation of enzymatic "sequencing engines" by single-molecule detection could substantially speed up and simplify sequencing by omitting amplification and cloning steps. Most fascinating is the fact that they would theoretically enable the reading of long DNA sequences with up to several thousand base pairs and accelerate the reading rates to hundreds of bases per second, depending on the enzyme involved. The concept of single-molecule SBS relies on observation of the stepwise incorporation of nucleotides into a growing chain of DNA and requires a technique for distinguishing the four monomers A, G, C, and T, preferentially using spectrally separable fluorescent dyes.

Although proposed as early as 20 years ago,[4] single-molecule sequencing has been difficult to realize because consistent, gapless sequence analysis would necessitate complete replacement of every nucleotide by its fluorescent analogue. However, protocols utilizing nucleotides with fluorescent dyes linked to the nucleobases are limited by increased steric hindrance during enzymatic incorporation, while approaches involving nucleotides with cleavable fluorescent dyes are hampered by repetitive cleaving and washing steps. Further major difficulties of single-molecule sequencing concern the isolation and handling of single target DNA molecules and the reliable spectroscopic identification of fluorescent labels against the background signal.

A recent breakthrough in single-molecule sequencing by synthesis was achieved by Korlach and Turner together with their team at Pacific Biosciences.[5] They were able to present a proof of concept for real-time, single-molecule sequencing by a DNA polymerase that performs uninterrupted, template-directed synthesis using four nucleotides linked to distinguishable fluorescent dyes. Two sophisticated techniques were crucial for their success: the use of zero-mode waveguide nanostructure arrays and covalent conjugation of fluorescent dyes to the 5′-phosphate moiety of nucleotides.

Conventional methods for studying single molecules, such as confocal or total internal reflection microscopy, provide observation volumes in the femtoliter range (10^{-15} L) and thus require pico- to nanomolar fluorophore concentrations to isolate individual molecules in solution. Polymerases (and many other enzymes), however, require substrate concentrations in the micromolar range, which in turn calls for a 1000-fold reduction of the conventional observation volume. Korlach approached this problem by developing nanophotonic structures, so-called zero-mode waveguides (ZMW), that consist a metal film with holes smaller than the light wavelength used deposited in arrays on a fused silica coverslip.[6] Below a critical waveguide size that is related to the wavelength of the light as well as to the waveguide shape, no propagating modes exist within the waveguide (= zero-mode waveguide). If illuminated by a laser from below, light intensity along such a ZMW decays exponentially. The resulting observation volume is then reduced to the atto- to zeptoliter range (10^{-18} to 10^{-21} L). In the case of the nucleotide concentrations of 0.1 to 10 μM that are necessary for fast, accurate, and processive DNA synthesis by a DNA polymerase, the average molecular occupancy of a ZMW with 100 nm diameter is roughly 0.01 to 1 molecule.

The problems associated with consecutive incorporation of base-linked fluorescent nucleotides were avoided by Korlach and Turner with the development and use of four

[*] J. Ibach, Priv.-Doz. Dr. S. Brakmann
Department of Chemical Biology/BCMT
Technische Universität Dortmund
Otto-Hahn-Strasse 6, 44227 Dortmund (Germany)
Fax: (+49) 231-9742-6627
E-mail: susanne.brakmann@tu-dortmund.de

nucleotides linked to fluorophores through a 5′-oligophosphate (Figure 1). These modified nucleotides acted as substrates for bacteriophage Φ29 DNA polymerase, which released the fluorophore from each incorporated nucleotide during phosphodiester bond formation, thus generating natural, unmodified DNA. Φ29 DNA polymerase in its wild-type form exhibits the highest processivity and strand displacement activity among known DNA polymerases and is capable of synthesizing DNA stretches with more than 70000 bases. With respect to the incorporation of phospholinked fluorescent nucleotides, optimization of both enzyme and linker structure was necessary. Therefore, Korlach and Turner introduced site-specific mutations into the polymerase and adjusted the linker length and chemistry of their nucleotides to allow for complete replacement of the natural nucleotides.

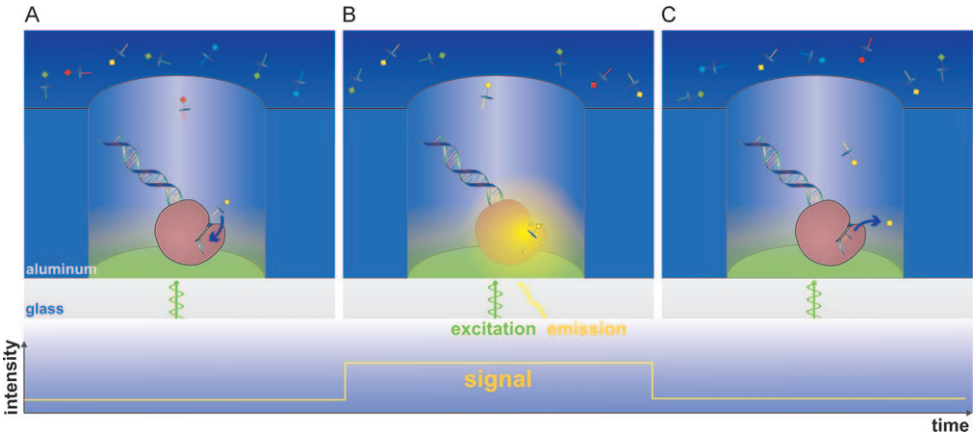

Figure 2. Sequence of events during incorporation of a single phospholinked nucleotide (according to Eid et al.[5]). The ZMW nanostructure reduces the excitation volume to 10^{-21} L, allowing for detection of single fluorescent substrate molecules against background signal originating from the bulk solution. A) Approach of a phospholinked nucleotide to its complementary template base. B) Cognate association within the polymerase active site causes an elevation of fluorescence emission on the corresponding color channel. C) Phosphodiester bond formation releases the dye–linker–oligo-/pyrophosphate, which diffuses out of the ZMW. Fluorescence emission drops to background signal.

In combination, these exciting developments enabled the exemplary reading of sequence data in real time. The experimental principle relies on the emission of a fluorescence pulse that is produced whenever the polymerase retains a cognate, color-coded nucleotide in the detection volume of the ZMW (Figure 2). Since duration of fluorophore retention is much longer than the time scales associated with diffusion (microseconds) or binding of a noncognate monomer (less than 1 ms), "signals" can be reliably distinguished from the low and constant background fluorescence.

Using a 150-base linear template, real-time four-color sequencing was demonstrated using a setup with two lasers for excitation of four fluorophores. Identification of fluorescence pulses took place after dye-weighted summation and automatic comparison of the measured spectra to four reference spectra by least-squares fitting. In a single run, the read that was extracted from measured pulses yielded 158 total bases, 131 of which were correctly identified, while 27 pulses were errors (12 deletions, 8 insertions, 7 mismatches). However, by averaging 15 sequencing runs, the median accuracy was improved to 99.3 % and thus to a value sufficient for resequencing applications. Detailed analysis showed that errors were attributed to 1) too short of intervals between two incorporation events (result: deletions), 2) dissociation of a cognate nucleotide from the polymerase active site before phosphodiester formation (result: insertions), and 3) misassignments arising from insufficient spectral separation (result: mismatches).

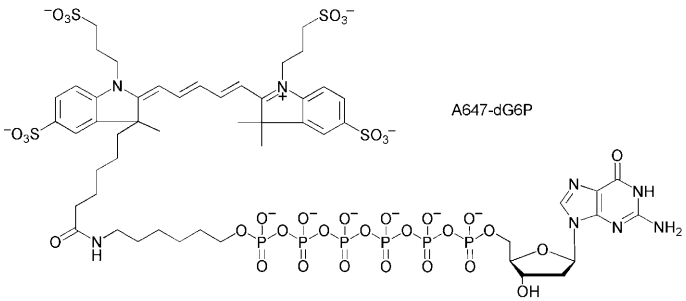

Figure 1. Example of a nucleotide labeled by linkage to the 5′-oligophosphate.[7] Cleavage of the label moiety by action of a DNA polymerase occurs between P_α and P_β.

The same analysis directly led to another important aspect of real-time monitoring at the single-molecule level: Through pulse width and interpulse duration, the system reports on the kinetics of every base incorporation and thereby resolves the distribution of kinetic data over hundreds of different sequence contexts. This aspect paves the way for unprecedented new information about DNA polymerases or, more generally, about enzyme kinetics.

Regarding future application of their technique for long-read sequencing, Korlach and Turner also investigated the potential of Φ29 DNA polymerase in rolling circle amplification with phospholinked fluorescent nucleotides. Their data showed that enzyme activity continued for more than one hour, during which more than 4000 nucleotides were incorporated. Consequently, even approximately 30000 functioning ZMWs would be sufficient for producing a raw read of the human genome per day (one-fold coverage): low-cost genome-scale sequencing is in sight. Thus, we are excited about the further evolution of this attractive field of research!

Received: January 17, 2009
Published online: May 4, 2009

[1] a) J. Shendure, R. D. Mitra, C. Varma, G. M. Church, *Nat. Rev. Genet.* **2004**, *5*, 335; b) M. Ronaghi, M. Uhlen, P. Nyrén, *Science* **1998**, *281*, 363.
[2] See: http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-003.html.
[3] M. L. Metzker, *Genome Res.* **2005**, *15*, 1767.
[4] J. H. Jett, R. A. Keller, J. C. Martin, B. L. Marrone, R. K. Moyzis, R. L. Ratliff, N. K. Seitzinger, E. Brooks-Shera, C. C. Stewart, *J. Biomol. Struct. Dyn.* **1989**, *7*, 301.
[5] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. deWinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, *Science* **2009**, *323*, 133.
[6] M. J. Levene, J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, W. W. Webb, *Science* **2003**, *299*, 682.
[7] J. Korlach, A. Bibillo, J. Wegener, P. Peluso, T. T. Pham, I. Park, S. Clark, G. A. Otto, S. W. Turner, *Nucleosides Nucleotides Nucleic Acids* **2008**, *27*, 1072.