
Gordon C. K. Roberts
Editor

Encyclopedia of Biophysics

With 1597 Figures and 131 Tables



Editor

Gordon C. K. Roberts
Honorary Professor of Biochemistry
Department of Biochemistry
University of Leicester
Leicester, UK

ISBN 978-3-642-16711-9 ISBN 978-3-642-16712-6 (eBook)

ISBN 978-3-642-16713-3 (print and electronic bundle)

DOI 10.1007/978-3-642-16712-6

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012949366

First Edition: Copyright European Biophysical Societies' Association (EBSA). English edition published by Springer-Verlag Berlin Heidelberg 2013. All rights reserved.

© European Biophysical Societies' Association (EBSA) 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher nor EBSA can accept any legal responsibility for any errors or omissions that may be made. The publisher and EBSA make no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Single-Molecule Sequencing

Jonas Korlach and Stephen W. Turner
Pacific Biosciences, Menlo Park, CA, USA

Definition

Single-molecule sequencing refers to techniques that can read the base sequence directly from individual strands of DNA or RNA present in a sample of interest.

Basic Characteristics

Nucleic acid sequencing represents one of the most important techniques in the study of biological systems – the order of the four different bases in the genome and transcriptome define the genetic blueprint of an organism. In addition to this basic hereditary genetic information, DNA also contains epigenetic modifications that are variable with respect to space (the organism's different tissues and cell types) and time (different stages in the organism's life cycle). While sequencing has already altered the way biological research is undertaken, extensive and cost-effective DNA and RNA sequencing, when combined through powerful computational methods with proteomics, biomolecular dynamics, and a host of other phenotypic information, is expected to precipitate a radical expansion of our understanding of biology, and to improve human health, food and energy supplies, and environmental conservation.

Decoding the sequence of bases in a *population* of identical DNA molecules became technically feasible in the late 1960s (reviewed in (Wu 1993)). These ensemble (non-single-molecule) sequencing technologies have steadily grown to provide ever greater DNA sequencing throughput at reduced cost (reviewed in [Mardis 2008; Shendure and Ji 2008]). Concurrently, there has also been great interest in pushing the DNA input sensitivity to the ultimate analytical limit – obtaining the sequence from an *individual* DNA molecule. The principal driving forces behind these efforts are:

1. *Increasing intrinsic sequencing speeds.* Ensemble DNA sequencing technologies rely on precisely

gating the identical processing of many DNA molecules during sequencing to assure synchronicity within the clonal DNA population. This is commonly achieved by terminating various enzymatic activities (such as DNA polymerization or ligation) after one or more bases, interrogation of the result, removal of the termination agent, and repetition of this cycle. The multi-step nature of this protocol (often including several washing steps) results in relatively slow cycle times and thus low intrinsic DNA sequencing speeds – often many minutes per base sequenced.

2. *Increasing the length of DNA sequencing reads.* In ensemble-based sequencing techniques, the DNA sequence quality decreases with sequence length because the yield for the gated addition of each base is less than 100%, thereby resulting in the population of molecules becoming less synchronized with each processing cycle. This limits the readlength of these techniques to a range of only several hundred bases. Single-molecule sequencing interrogates each DNA molecule independently, thereby avoiding this problem of loss of synchrony or *dephasing*, resulting in longer readlengths with uniform sequence quality.
3. *Avoiding processing steps prior to sequencing.* Ensemble DNA sequencing technologies routinely require amplification techniques, such as polymerase chain reaction (PCR) or whole-genome amplification (WGA). These steps have the potential to introduce limitations and errors, e.g., limiting the size of DNA fragments, introducing population bias through uneven amplification efficiencies, and base sequence errors introduced during the amplification process. They also increase the overall time and cost of sequencing.
4. *Resolving heterogeneous DNA mixtures.* Commonly, natural samples contain a mixture of different DNA molecules, e.g., maternally and paternally inherited alleles, environmental microbial samples, tissue biopsies containing both healthy and tumor cells, or forensic samples from crime scenes. To address this problem, some ensemble-based sequencing techniques employ sample preparation steps that proceed through dilution and subsequent amplification steps that produce clonal populations (originating from a singly diluted molecule) for sequencing. In addition to the concerns around

Single-Molecule Sequencing, Table 1 Commercialization of different single-molecule sequencing methods

Company	Method	Website
Halcyon Molecular	Transmission electron microscopy (TEM)-based SMS	www.halcyonmolecular.com
Helicos Biosciences	True Single-Molecule Sequencing (tSMS™)	www.helicosbio.com
IBM	Transistor-based nanopore DNA sequencing	www.ibm.com
Life Technologies	Qdot-based SMS	www.lifetechnologies.com
NobleGen Biosciences	Optical readout nanopore DNA sequencing	www.noblegenbio.com
Oxford Nanopore Technologies	Nanopore DNA sequencing	www.nanoporetech.com
Pacific Biosciences	Single-Molecule, Real-time (SMRT™) DNA sequencing	www.pacificbiosciences.com
Reveo	Scanning tunneling microscopy (STM)-based SMS	www.reveo.com
ZS Genetics	Transmission electron microscopy (TEM)-based SMS	www.zsgenetics.com

amplification described in (3), these steps typically break the original DNA fragments into small pieces, thereby confounding the task of correctly assembling the native constitution of the heterogeneous DNA sample.

5. *Reading DNA base modifications as part of sequencing.* Base modifications are present in the genomes of virtually all organisms, as well as viruses and phages. They include epigenetic markers essential for growth and development (5-methylcytosine (5-mC), 5-hydroxymethylcytosine (5-hmC)), bacterial identity marks important in host-phage interactions (5-mC, 5-hmC, glucosylated 5-hmC, 6-methyladenine, 4-methylcytosine), or products of DNA damage (e.g., 8-oxoguanine, BPDE adducts (a major form of damage from cigarette smoke), thymine dimers) (Murray 2002; Siede et al. 2005; Tollefsbol 2010). Currently, the vast majority of these modifications cannot be sequenced with ensemble-based sequencing techniques, and where possible require complex chemical steps (e.g., bisulfite conversion of 5-methylcytosine). Single-molecule sequencing technologies aim to read these important DNA marks as an integral part of the sequencing method.

Several different single-molecule DNA sequencing techniques have been developed or are currently in development (reviewed in [Xu et al. 2009; Schadt et al. 2010b]), and are being commercialized (Table 1).

There are currently two principal approaches: (1) DNA polymerase-based single-molecule DNA sequencing, and (2) direct inspection of individual DNA strands.

DNA Polymerase-Based Single-Molecule Sequencing

Gated Single-Molecule Sequencing-By-Synthesis

Gated sequencing-by-synthesis – at the heart of many ensemble-based DNA sequencing approaches – has been adapted to the single-molecule regime in a technology commercialized by Helicos Biosciences. In its principle, individual DNA molecules are extended and detected one base at a time through the use of DNA polymerase and fluorescently labeled, terminating nucleotides that carry a cleavable moiety to remove the termination group after the detection step. The technology meets the desired performance criteria (3) and (4) described above – avoiding amplification prior to sequencing, and interrogating DNA molecules asynchronously, thereby allowing the sequencing of mixed samples. However, because sequencing is still based on gating the polymerase activity, it does not meet criteria (1) and (2), with cycle times of ~2 h per base, and average readlengths of ~35 bases.

Single-Molecule, Real-Time DNA Sequencing

Engineered by nature, DNA polymerases are arguably the most advanced sequencing machines. Over millions of years of molecular evolution, they have been optimized to rapidly and faithfully replicate genomes. Their speed, accuracy, processivity, and sensitivity to base modifications can be harnessed directly in single-molecule, real-time sequencing approaches using DNA polymerase – not as a reagent consumable, but as the sequencing engine itself.

Single-molecule sequencing by monitoring such *uninterrupted* activity of DNA polymerase is realized

in Single-Molecule, Real-Time (SMRT™) DNA sequencing commercialized by Pacific Biosciences. Phospholinked nucleotides – carrying the fluorescent label attached to the terminal phosphate moiety – are employed to allow continuous DNA synthesis, as the fluorescent label is removed by the natural enzymatic activity of the DNA polymerase during base incorporation. The enzyme/DNA complex is placed in a photonic nanostructure called a zero-mode waveguide (ZMW), which confines the observation volume to the immediate vicinity of the enzyme, thereby making the polymerase activity visible against the solution background of diffusing labeled nucleotides, and making gating and washing steps unnecessary. Individual DNA molecules are independently and asynchronously sequenced on arrays containing tens of thousands of ZMW, with intrinsic sequencing speeds of >1 base per second, and readlength averaging >1,000 bases. On topologically circular DNA template molecules, the polymerase will undergo rolling-circle DNA synthesis, allowing a consensus sequence to be obtained from multiple passes on the same molecule (Travers et al. 2010). Using this circular consensus sequencing approach, the error-free sequencing of an individual DNA molecule greater than 1 kb in length (segment 7 of influenza strain H1N1 (A/New York/1682/2009); 1,027 bp) has been reported (Schadt et al. 2010a). The polymerase kinetics is affected by the DNA template structure, allowing the sequencing and identification of epigenetic markers and other DNA base modifications (Flusberg et al. 2010).

Another real-time sequencing approach using quantum dots for labeling polymerase molecules and nucleotides in conjunction with fluorescence resonance energy transfer as an alternative observation volume confinement strategy is pursued by Life Technologies, but little information is available about the feasibility of this approach.

Direct Inspection of Individual DNA Strands

Nanopore-Based Single-Molecule Sequencing

In nanopore-based single-molecule sequencing approaches, DNA molecules are individually driven through nanoscale pores of dimensions that only permit the passage of single-stranded DNA in strict linear sequence. Nanopores can be of biological origin – membrane protein complexes such as α -hemolysin or *Mycobacterium smegmatis* porin A (MspA) protein – or they can be created

synthetically from solid state materials, such as silicon nitride, aluminum oxide, multilayer graphene sandwiches, or carbon nanotubes.

The pore contains a detection region capable of differentiating the four different bases, thereby generating a temporal signal corresponding to the sequence of bases as they traverse the pore. Different types of signals are being evaluated, including differential ionic current blockades through the pore, tunneling currents across the pore, or capacitance changes. Using an α -hemolysin protein nanopore with a covalently attached cyclodextrin adapter molecule as the detector reading head, the four bases and 5-methylcytosine could be distinguished as free nucleotide monophosphates by measuring the residual pore current (Clarke et al. 2009).

Precise control over the DNA transport through the pore is crucial to the success of these methods, and has in some instances been implemented by the use of enzymatic activities, such as polymerases or exonucleases positioned close to the pore's entrance. Using this approach, the real-time detection of several sequential nucleotide additions by a DNA polymerase has been measured by differential current blockage of an α -hemolysin pore as the DNA template strand was drawn through the nanopore lumen during replication (Chu et al. 2010; Lieberman et al. 2010). To address potential challenges related to achieving large multiplex factors of independently monitoring the small electrical signals from each nanopore, one alternative strategy has been reported which allows for the parallel optical readout from multiple nanopores (McNally et al. 2010).

Microscopy-Based Single-Molecule Sequencing

Advanced electron microscopy techniques are being pursued to directly image the sequence of bases in a DNA template molecule. The different approaches include transmission electron microscopy (TEM) for direct chemical detection of atoms that would uniquely identify the different bases. Other approaches include prior differential chemical labeling of the bases with compounds such as iodine and bromine to enhance detectability and contrast. A third approach encompasses the imaging of single-stranded DNA molecules by scanning tunneling microscopy (STM), and detection of the four bases by differential tunneling current between the atomically sharp STM tips, the base, and a conductive surface. Using this approach, an

“electronic fingerprint” of guanine bases in single-stranded DNA molecules of known sequence has been demonstrated (Tanaka and Kawai 2009).

Cross-References

- [DNA Polymerase](#)
- [Electron Microscopy](#)
- [Single-Molecule Fluorescence Resonance Energy Transfer](#)
- [Single-Molecule Methods](#)
- [Zero-Mode Waveguides](#)

References

- Chu J, Gonzalez-Lopez M, Cockroft SL, Amarin M, Ghadiri MR. Real-time monitoring of DNA Polymerase function and step-wise single-nucleotide DNA strand translocation through a protein nanopore. *Angew Chem Int Ed Engl*. 2010;49:10106–9.
- Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol*. 2009;4:265–70.
- Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Meth*. 2010;7:461–5.
- Lieberman KR, Cherf GM, Doody MJ, Olasagasti F, Kolodji Y, Akeson M. Processive Replication of Single DNA Molecules in a Nanopore Catalyzed by phi29 DNA Polymerase. *J Am Chem Soc*. 2010;132:17961–17972.
- Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*. 2008;9:387–402.
- McNally B, Singer A, Yu Z, Sun Y, Weng Z, Meller A. Optical recognition of converted DNA nucleotides for single-molecule DNA sequencing using nanopore arrays. *Nano Lett*. 2010;10:2237–44.
- Murray NE. Immigration control of DNA in bacteria: self versus non-self. *Microbiology*. 2002;148:3–20.
- Schadt EE, Korlach J, Turner S. Presentation at the American Society of Human Genetics 60th annual meeting, Pacific Biosciences commercial workshop, Washington, DC, 2–6 Nov 2010a
- Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. *Hum Mol Genet*. 2010b;19:R227–40.
- Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*. 2008;26:1135–45.
- Siede W, Kow YW, Doetsch PW. DNA damage recognition. Boca Raton: Taylor & Francis; 2005.
- Tanaka H, Kawai T. Partial sequencing of a single DNA molecule with a scanning tunnelling microscope. *Nat Nanotechnol*. 2009;4:518–22.
- Tollefsbol TO. Handbook of epigenetics: the new molecular and medical genetics. 1st ed. Amsterdam: Academic; 2010.
- Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res*. 2010;38:e159.
- Wu R. Development of enzyme-based methods for DNA sequence analysis and their applications in the genome projects. *Adv Enzymol Relat Areas Mol Biol*. 1993;67:431–68.
- Xu M, Fujita D, Hanagata N. Perspectives and challenges of emerging single-molecule DNA sequencing technologies. *Small*. 2009;5:2638–49.

Single-Molecule Spectroscopy

Benjamin Schuler

University of Zurich, Switzerland

Introduction

We are used to conceptualizing chemical reactions and biochemical processes in terms of individual molecules. We do so on various levels of complexity, ranging from simple textbook cartoons to the atomic detail of ► [molecular dynamics simulations](#) (► [Molecular Dynamics Simulations of Lipids](#)). But, remarkably, the vast majority of our knowledge about these systems has been derived from experiments on large ensembles of molecules, which yield only average values of observable properties. The sequence of molecular events describing the underlying reactions is typically inferred from testing models by systematic variation of parameters. For kinetic ensemble studies, the reactions need to be synchronized, which is often difficult. The concept of observing single molecules is particularly appealing for processes with a large degree of conformational and dynamic heterogeneity, which abound in biological systems.

Only rather recently has it become feasible to investigate biological systems on the level of single molecules with spectroscopic methods. These techniques offer a fundamental advantage beyond our mere fascination for the direct observation of molecular processes: they can resolve and quantify the properties of individual molecules or subpopulations, which are often inaccessible in ensemble experiments. ► [Fluorescence spectroscopy](#) (► [Fluorescence and FRET in Membranes](#)) has been a particularly popular technique, owing to its extreme sensitivity and versatility. For example, single-molecule spectroscopy