Carbon 115 (2017) 293–311

Contents lists available at ScienceDirect

Carbon

journal homepage: www.elsevier.com/locate/carbon

Review article

Nanocarbons for DNA sequencing: A review

Nianjun Yang*, Xin Jiang**

Institute of Materials Engineering, University of Siegen, Siegen, 57076, Germany

ARTICLE INFO

Article history:
Received 2 November 2016
Received in revised form
4 January 2017
Accepted 6 January 2017
Available online 9 January 2017

Keywords:
DNA sequencing
Nanopores
Graphene
Carbon nanotubes
Diamond

ABSTRACT

DNA sequencing using nanopores, so-named nanopore sequencing, is one of the most promising and revolutionary DNA sequencing technologies to provide the order and the type of nucleobases in a DNA molecule. In compare to those existing materials (e.g., Si_3N_4, polymer, etc.), nanocarbons (e.g., graphene, carbon nanotubes, etc.) are perfect materials for nanopore sequencing. This article provides an overview on the general issues of nanopore sequencing and summarizes recent progress and achievements of nanopore sequencing using graphene and carbon nanotubes. The future research directions using nanocarbons for nanopore sequencing are discussed and outlined.

Contents

* Corresponding author.
** Corresponding author.
E-mail addresses: nianjun.yang@uni-siegen.de (N. Yang), xin.jiang@uni-siegen.de (X. Jiang).

http://dx.doi.org/10.1016/j.carbon.2017.01.012
0008-6223/

1. Introduction

A deoxyribonucleic acid or a DNA molecule consists of a monosaccharide sugar (so-called deoxyribose), a phosphate group, and four kinds of nitrogen-containing nucleobases, namely, cytosine (C), guanine (G), adenine (A), and thymine (T). The type and the order of these nucleobases in a DNA molecule actually carry a broad range of biological information and genetic instructions at the molecular level. Such a DNA sequence thus provides the blueprint of our life [1,2]. Therefore, it is crucial to visualize these sequences, so-called DNA sequencing. From obtained DNA sequences, genomic data can be collected [3,4]. In this way, early diagnosis and treatment as well as understanding the mechanisms of genetically related diseases can be conducted. The side effects from delivered drugs can be avoided and personalized genomic medicines can be developed. Eventually, the prevention of those diseases can be realized [1,4].

There is already a rich and diverse history to develop revolutionary, fast, reliable, and inexpensive DNA sequencers [5–8]. In 2004 the National Human Genome Research Institute of the National Institutes of Health, USA launched a so-called "$1000 Genome" project. Within this project, four Gold Standards have been proposed for DNA sequencing: high accuracy (<1 error/10,000 bases), no gaps for the long read length (>10 Kb or longer), high throughput with a short turnaround time per run (in the matter of hours or even minutes), and low cost (<$1000/genome) [5–8]. Note here that the price only includes the usage and analysis costs. The cost for the disposal and large instrument is not covered. To reach above four Gold Standards for DNA sequencing, numerous DNA sequencing technologies have been developed over past years [5–8].

We classify in this paper the reported DNA sequencing methods into four generations [1]:

Sanger sequencing as the first generation: in such a process DNA strands are synthesized starting from a known primer sequence and terminated by a specific dideoxy deoxyribonucleoside triphosphate (dNTP). In other words, the last base in the sequence is known. DNA strands are then size-separated by gel electrophoresis for reading off the last base. The Sanger procedure is time-consuming and only has small throughput due to slow DNA fragment separation in gels. Moreover, Sanger sequencing is an expensive process. Sequencing of 100 genes from 100 samples with a non-commercial or commercial sequence service needs $300,000 to $1,000,000. It is thus only useful for small-scale projects in the kilobase (kb)-to-megabase (mb) range. However, this technique actually provides the gold standards (e.g., accuracy) for DNA sequencing [5].

Amplification-based massively parallel sequencing as the second generation: it involves *in vitro* amplification of DNA strands and their clustering onto dedicated surfaces as well as the sequencing via synthesizing the arrays of micro-beads, DNA nano-balls, and DNA clusters, etc. For example, fluorescently tagged nucleotides have been added by a polymerase, enabling a signal for each base to be instantly read off. These improvements substantially increase the degree of parallelism and reduce reagent volumes, leading to much faster and cheaper sequencing. These methods come at the cost of significantly lower read lengths (typically ~100 bp) than the Sanger method (>500 bp) [1].

Single-molecule sequencing as the third generation: it includes stepwise and real-time single-molecule sequencing by synthesis, single-molecule motion sequencing, Raman scattering based sequencing, electron microscopy, molecular force spectrometry, and so on [5,9]. Unfortunately, they suffer from complicated sample preparation, challenging algorithms for data processing, low throughput, short read lengths, and high cost [1].

Nanopore sequencing as the fourth generation: it is a nanopore-based DNA sequencing. Since nanopore sequencing promises inexpensive, reliable, high-throughput sequencing, and finally accomplishes above Gold Standards, it is trusted to bring genomic science into personal medicine [1,6,10–17]. In past years, lots of progress and achievements have been thus achieved. Several companies, such as Oxford Nanopore Technologies, Life, and Roche with IBM, Electronic BioSciences, NABsys, and Genia have been extensively involved with respect to nanopore sequencing. Up to date, a few nanopore sequencers are already market-available. For example, in 2012 Oxford Nanopore Technologies released one MiniION only with a price of $900) [18].

In this review, we present first several general issues related with nanopore sequencing, covering its history, sequencing mechanisms, employed nanopores, and existing challenges. In the second session, recent progress and achievements of nanopore sequencing by use of graphene nanostructures for DNA sequencing are summarized. The fabrication and properties of different graphene nanostructures (e.g., nanopore, nanogap, etc.) are presented as well. In the third part, DNA sequencing using single-walled and multi-walled carbon nanotubes (CNTs) is shown. Before conclusion remarks, the future directions of nanocarbon based nanopore sequencing, including the development of novel sequencing techniques and the employment of new carbon materials, are discussed and outlined. The employment of diamond for nanopore sequencing is highlighted. We hope this article is informative and useful for material scientists, chemists, biologists, and engineers who are devoted themselves to nanopore sequencing as well as to the research activities of carbon materials.

2. General issues of nanopore sequencing

2.1. History

Deamer, Church, Kasianowicz, and Branton first envisioned the idea of nanopore sequencing in the 1990s [19]. In 1996, Deamer, Branton and their colleagues demonstrated for the first time the