Analytica Chimica Acta 852 (2014) 274–283

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

A real-time decoding sequencing based on dual mononucleotide addition for cyclic synthesis

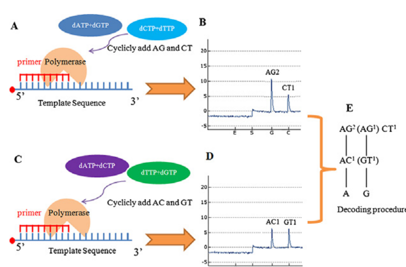
Dan Pu [a], Yuhua Qi [b,1], Lunbiao Cui [b], Pengfeng Xiao [a,*], Zuhong Lu [a,*]

[a] State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China
[b] Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, Jiangsu 210009, China

HIGHLIGHTS

- Templates are determined without directly measuring the base sequence in this method.
- Templates are sequenced with the incorporation of AG/CT, AC/GT or AT/CG.
- Templates will be sequentially decoded by two sets of encodings.
- This method applies fewer cycles to obtain longer read length.
- This method is able to be applied to differentiate nucleic acid sequences.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 2 June 2014
Received in revised form 28 August 2014
Accepted 8 September 2014
Available online 16 September 2014

Keywords:
Sequencing by synthesis (SBS)
Pyrosequencing
Dual mononucleotide addition
Decoding
Encoding
Species differentiation

ABSTRACT

We propose a real-time decoding sequencing strategy in which a template is determined without directly measuring base sequence but by decoding two sets of encodings obtained from two parallel sequencing runs. This strategy relies on adding a mixture of different two-base pair, A + G, C + T, A + C, G + T, A + T or C + G (abbreviated as AG, CT, AC, GT, AT, or CG), into the reaction each time. When a template is cyclically interrogated twice with any two kinds of dual mononucleotide addition (AG/CT, AC/GT, and AT/CG), two sets of encodings are obtained sequentially. The two sets of encodings allow for the bases to be sequentially decoded, moving from first to last, in a deterministic manner. This strategy applies fewer cycles to obtain longer read length compared to the traditional real-time sequencing strategy [1]. Partial *rnpB* gene was applied to verify the applicability of the decoding strategy via pyrosequencing. The results indicated that the sequence could be reconstructed by decoding two sets of encodings. Moreover, streptococcal strains could be differentiated by comparing signal intensity in each cycle and encoding size of each template. This strategy is likely to be applied to differentiate nucleic acid sequence as encoding size and signal intensity in each cycle vary with the base size and composition. Furthermore, it has the potential in building a promising strategy that could be utilized as an alternative to conventional sequencing systems.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Primer-directed polymerase extension is able to incorporate thousands of base pairs, which indicates sequencing-by-synthesis (SBS) based technologies have great potential in read length. The existing SBS based high-throughput sequencing methodologies are

* Corresponding author. Tel.: +86 25 83793310; fax: +86 25 83793310.
E-mail addresses: xiaopf@seu.edu.cn (P. Xiao), zhlu@seu.edu.cn (Z. Lu).
[1] The author contributed equally to this work.

http://dx.doi.org/10.1016/j.aca.2014.09.009
0003-2670/© 2014 Elsevier B.V. All rights reserved.

classified as single nucleotide addition [2–5], and four nucleotides addition [6–10]. The former is to add only one kind of nucleotide into a reaction to determine the number of incorporated nucleotides; the latter is to add special modified monomers to determine the type of incorporated nucleotides. In high-throughput DNA sequencing, the length of one read is an important indicator to measure the sequencing strategies. As for SBS, the read length is relevant to the number of reaction steps as well as the type of incorporated nucleotides. The large number of steps to read one base pair lends itself to significant inefficiencies [11]. In SBS with four nucleotides addition, a function of the read length and the cycle efficiency (Ceff) is as following: $(\text{Ceff})^{\text{readlength}} = 0.5$ [12]. When Ceff is 90%, the read length is 7 bp. However, when Ceff is more than 99%, the read length is greater than 100 bp. Therefore, an additional step (such as cleavage or incorporation of nucleotides) is likely to affect the Ceff, and eventually affect read length. In addition, the use of modified nucleotides gives rise to much shorter read lengths because of asynchrony [13]. That is probably why the read length from the existing commercialized sequencing instruments which use natural nucleotides as raw materials is longer than those using modified nucleotides.

Real-time sequencing methods have many advantages, such as maintaining the characteristics of natural nucleotides and eliminating the subsequent processing in the next sequencing cycles, but there is still one drawback that not every sequencing reaction gives an efficient message which may affect the Ceff, thus affecting the read length. There are strategies that may extend the read length by use of fewer than four labeled nucleotides. One of these strategies utilizes six different runs of different two-base pair combinations. The sequence is reconstructed from the order of each of the two bases [14]. Another strategy, called the ordered label strategy, relies on using all combinations of two labeled nucleotides to determine the order for each set, and then reconstructing the full sequence information [15]. Not long ago, some DNA sequencing technologies (ABI SOLiD sequencing) have been developed, which do not directly measure the base sequence, but measure DNA bases in pairs and in an encoded form, such as two-base encoding. In order to compare such data to a reference sequence, the encodings must be decoded into base sequence [16,17]. This technology has great potential in error tolerance by differentiating biological variants from sequencing errors. However, read length is limited for fluorescently-labeled nucleotides and for large number of steps to read one base pair. In this report, we addressed a real-time decoding sequencing which contained the step of sequencing template sequence twice with dual mononucleotide addition, obtaining two sets of encodings, and finally reestablishing the base sequence with the two sets of encodings. This strategy relied on adding one of different two-base pair combinations, AG, CT, AC, GT, AT and CG, into the reaction each time. Here, AG, CT, AC, GT, AT and CG referred to reagent additions containing mixtures of those mononucleotide triphosphates. These different two-base pair combinations could form three kinds of dual mononucleotide addition, AG/CT, AC/GT, and AT/CG. This strategy interrogated the template sequence by any two combinations of AG/CT, AC/GT, and AT/CG. It is based on the principle that the signal intensities of released detection molecules are proportional to the number of incorporated nucleotides. When dual mononucleotide are added into the reaction each time, much stronger signal intensities are captured, making the detection of a trace of template possible. Moreover, the encodings obtained from a single sequencing run is also found to be useful for differentiation of nucleic acid sequence, since the signal intensity in each cycle and encoding size vary with the sequence composition and size. This strategy applies fewer cycles to obtain longer read length compared with traditional real-time sequencing strategy. It is compatible with most of the commercial sequencing instruments and is likely to be used as an alternative to the sequencing system. We hope it will provide the researchers with a new technology to analyze nucleic acid sequence.

2. Materials and methods

2.1. Sequences and reagents

All the oligonucleotide sequences used are shown in Table 1. Synthetic sequences were purchased from Invitrogen (Shanghai, China). 5′ modifications were also performed by Invitrogen™ (Shanghai, China). The SQA PyroMark Gold Q96 Reagents (1 × 96), and solutions including annealing buffer (20 mM Tris–Acetat, 2 mM $MgAc_2$, pH 7.6), denaturation buffer (0.2 M NaOH), wash buffer (10 mM Tris–Acetate, pH 7.6) and binding buffer (10 mM Tris–HCl, 2 M NaCl, 1 mM EDTA, 0.1% Tween 20, pH 7.6) were purchased from QIAGEN (Hilden, Germany).

2.2. Pyrosequencing with one nucleotide addition

Pyrosequencing with one nucleotide addition was performed on single-stranded synthetic templates. 1 μM synthetic template (T1) and 1.2 μM sequencing primer (SP1) were added into a 96-well filter plate and the liquid was annealed by incubating in the presence of annealing buffer for 2 min at 80 °C, and then cooled to room temperature for use.

Pyrosequencing was carried out on an automated PSQ 96MA system (Biotage AB, Uppsala, Sweden) by using the SQA PyroMark Gold Q96 Reagents (1 × 96), containing enzyme, substrate, and nucleotides. The reaction cascaded primarily by cyclic supplying of A/C/G/T into the growing DNA chains. The pattern of emitted light in relation to the nucleotide dispensation order and number of nucleotides incorporated was subsequently illustrated on a pyrogram. The data were then analyzed by the pyrosequencing software.

2.3. The decoding sequencing via pyrosequencing

Pyrosequencing with dual mononucleotide addition was first performed with a set of single-stranded synthetic templates to investigate whether the signal intensities of released pyrophosphates were proportional to the number of incorporated nucleotides. 1 μM synthetic template (T2–T7) and 1.2 μM sequencing primer (SP1) were annealed and cooled to room temperature. This strategy included two parallel sequencing runs. In the first sequencing run, pyrosequencing reactions cascaded primarily by cyclic supplying of AG/CT into the growing DNA chains. A set of encodings were obtained. In the second sequencing run, the

Table 1
Sequences of synthesized oligonucleotides used in this study.

Template	Sequence (5′-3′)
T1	CGCTTTCCTCTCTATGGGCAGTCGGTGATCTAGTAGCATCGAGAC TAGGTGCACTGCTGTACCGTACATCCGCCTTGGCCG
T2	CGCTTCCTCCCCCCCCCCCACTGTACCGTACATCCGCCTTGGCCG
T3	CGCTTCCTCCCCCCCCCAAACTGTACCGTACATCCGCCTTGGCCG
T4	CGCTTCCTCCCCCCCAAAAACTGTACCGTACATCCGCCTTGGCCG
T5	CGCTTCCTCCCCCAAAAAAACTGTACCGTACATCCGCCTTGGCCG
T6	CGCTTCCTCCCAAAAAAAAACTGTACCGTACATCCGCCTTGGCCG
T7	CGCTTCCTCAAAAAAAAAAAACTGTACCGTACATCCGCCTTGGCCG
SP1	GCCAAGGCGGATGTACGGTACAG
SP2	CAATTTTTGGATAATCG
Forward primer	GTGCAATTTTTGGATAATCG
Reverse primer	Biotin-TGGGTTGCTAGCTTGAGG

The underlined segments were the hybridization regions with the sequencing primers SP1. SP1 and SP2 were sequencing primers.

Download English Version:

https://daneshyari.com/en/article/7555673

Download Persian Version:

https://daneshyari.com/article/7555673

Daneshyari.com