生物物理 47（2），129-132（2007）

理論／実験 技術

Pyrosequencing Technology for Short DNA Sequencing and Whole Genome Sequencing

Baback Gharizadeh[1], Mehran Ghaderi[2] and Pål Nyrén[3]
[1]Stanford Genome Technology Center, Stanford University
[2]Clinical Pathology/Cytology, Karolinska University Hospital
[3]Department of Biotechnology, Royal Institute of Technology

1. Introduction

Recent impressive advances in DNA sequencing technologies have accelerated the detailed analysis of genomes from many organisms. We have been observing reports of complete or draft versions of the genome sequence of several well-studied, multicellular organisms. Human biology and medicine are in the midst of a revolution by Human Genome Project[1)] as the main catalyst.

The chain termination sequencing method, also known as Sanger sequencing, developed by Frederick Sanger and colleagues[2)], has been the most widely used sequencing method since its advent in 1977 and still is in use after more than 29 years since its advent. The remarkable advances in chemistry and automation to the Sanger sequencing method has made it to a simple and elegant technique, central to almost all past and current genome-sequencing projects of any significant scale. Despite all these grand advantages, there are limitations in this method, which could be complemented with other techniques.

Among the current state-of-art DNA sequencing techniques, Pyrosequencing[3)] has emerged, which is being used for a wide variety of applications. In the beginning, the method was only restricted to single nucleotide polymorphisms (SNP) genotyping[4), 5)] and short reads[6), 7)] when it was introduced in 1997 but it is now being used for broader applications.

2. Pyrosequencing chemistry

Pyrosequencing technique is based on sequencing-by-synthesis principle[8), 9)] and employs a series of four enzymes to accurately detect short nucleic acid sequences during DNA synthesis. In Pyrosequencing[10)], the sequencing primer is hybridized to a single-stranded DNA biotin-labeled template (post-PCR alkali treated) and mixed with the enzymes; DNA polymerase, ATP sulfurylase, luciferase and apyrase, and the substrates adenosine 5′ phosphosulfate (APS) and luciferin.

Fig. 1 illustrates schematically the chemistry of Pyrosequencing method. Cycles of four deoxynucleotide triphosphates (dNTPs) are separately added to the reaction mixture iteratively. After each nucleotide dispensation, DNA polymerase catalyzes the incorporation of complementary dNTP into the template strand. Each nucleotide incorporation event is followed by release of inorganic pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide. ATP sulfurylase quantitatively converts PPi to ATP in the presence of APS. The generated ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin, producing visible light in amounts that are proportional to the amount of ATP. The light in the luciferase-catalyzed reaction is then detected by a photon detection device such a charge coupled device (CCD) camera or photomultiplier. The generated light is observed as a peak signal in the pyrogram (corresponding to electropherogram in dideoxy sequencing). Each signal peak is

1分子微弱発光を利用したパイロシークエンス法による実時間ゲノム配列計測技術
Baback GHARIZADEH[1], Mehran GHADERI[2] and Pål NYRÉN[3]
[1]Stanford Genome Technology Center, Stanford University
[2]Clinical Pathology/Cytology, Karolinska University Hospital
[3]Department of Biotechnology, Royal Institute of Technology

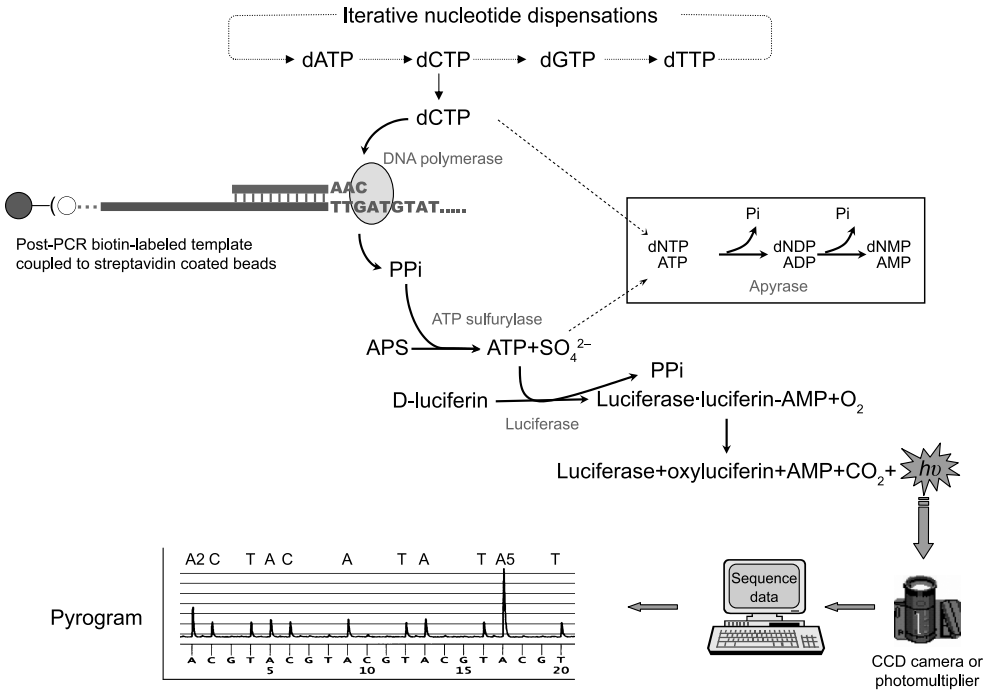

Fig. 1

Schematic illustration of Pyrosequencing method. Pyrosequencing is a non-electrophoretic real-time DNA sequencing method that uses the luciferase-luciferin light release as the detection signal for nucleotide incorporation into target DNA. The four different nucleotides are dispensed iteratively to a four-enzyme mixture. The pyrophosphate (PPi) released in the DNA polymerase-catalyzed reaction is quantitatively converted to ATP by ATP sulfurylase, which provides the energy to firefly luciferase to oxidize luciferin and generate light ($h\upsilon$). The light is detected by a photon detection device and monitored in real time by integrated software in a format called pyrogram. Finally, apyrase catalyzes degradation of nucleotides that are not incorporated and the sequencing reaction will be ready for the next nucleotide addition.

proportional to the number of nucleotides incorporated (a triple dGTP nucleotide incorporation generates a triple higher peak). Apyrase is a nucleotide-degrading enzyme, which continuously degrades ATP and non-incorporated dNTPs in the reaction mixture. There is a certain time interval (usually 65 seconds) between each nucleotide dispensation to allow complete degradation. For this reason, dNTP addition is performed one at a time.

During this synthesis process, the DNA strand is extended by complementary nucleotides, and the DNA sequence is demonstrated by the signal peaks in the pyrogram on a screen. Base-callings are performed with integrated software, which has many features for related SNP and sequencing analysis.

3. Applications of Pyrosequencing

Pyrosequnecing is well suited for *de novo* sequencing and resequencing. Pyrosequencing was earlier limited to sequencing of short stretches of DNA, due to the inhibition of apyrase. The natural dATP was a substrate for luciferase, resulting in false sequence signals. dATP was substituted by dATP-α-S[11]. Our experiments showed that higher concentrations of this nucleotide had inhibitory effect on apyrase catalytic activity, causing non-synchronized extension[12]. It was found out later that the dATP-α-S consisted of two isomers Sp and Rp. The Rp isomer was not incorporated in the DNA template as it was not a substrate for DNA polymerase, and its presence in the sequencing reaction simply inhibited apyrase activity. By introducing the dATP-α-S Sp isomer, substantial longer reads were achieved. This improvement had a major impact on Pyrosequencing read length and allowed sequencing of up to one hundred bases[12] and opened many avenues for numerous applications.

Currently, the Pyrosequencing method is broadly being applied for many applications such as SNP genotyping[4), 5), 13)] and for identification of bacteria[14)-16)], fungal[17)-19)] and viral typing[6), 20)-23)]. Moreover, the method has shown the ability for determination of difficult secondary structures[24)], mutation detection[25), 26)], DNA methylation analysis[27)-29)], multiplex sequencing[30), 31)], tag sequencing of cDNA library[7)] and clone checking[32)].

Another highly significant application is whole genome sequencing[33)]. Pyrosequencing is now being applied in microfluidic format commercially by 454 Life Sciences Corporation (Branford, CT, USA). The microfluidic pyroseqeuncing has been integrated

with emulsion PCR and DNA sequencing assembly software. The DNA sequence signal peaks are presented in flowgrams. Sequence reads of up to 100 bases are generated, allowing massive parallel fashion sequencing in picoliter volumes. The new platform has the capacity to sequence up to 300,000 samples and generate up to 20-40 million bases at an accuracy of 99% per 4 hours sequencing run.

In brief, the entire genomic DNA is sheared, and the fragments are ligated to adapters. The fragments are bound to the beads that favor one fragment per bead by limiting dilution approach. The beads are captured in the droplets of a PCR-reaction-mixture-in-oil[33]. The DNA fragments are clonally amplified and the emulsions are broken after complete amplification. After single strand treatments, the beads are deposited into the picotiter wells (etched in a fiber-optic slide). Each well has the capacity for one single bead. Smaller beads carrying immobilized enzymes for Pyrosequencing reaction are added to the wells and the regents flow cyclically through the wells. The generated photons are detected by a CCD camera and presented in flowgrams in integrated software.

4. Advantages and challenges

Pyrosequencing has emerged as an alternate method. Although it is relatively facing read-length limitations compared with dideoxy sequencing, it is a fast method with real-time read-out that is highly suitable for sequencing short stretches of DNA. Unlike Sanger sequencing, which lays a reading gap of roughly 20-30 bases from the sequencing primer, Pyrosequencing has the benefit that it can generate sequence signals immediately downstream of the primer. Sample preparation and single-strand DNA process is also relatively rapid (about 15 min) while sample preparation takes approximately 4 hours for Sanger sequencing (60 min for PCR cleanup, 3-4 hours for cyclic amplification and 15 minutes for dye cleanup). The reagent costs are considerably lower for sequencing short stretches of DNA compared to current available methods.

Homopolymer Ts (more than 3-4) are a challenge in Pyrosequencing. Homopolymer strings (mainly homopolymeric T) regions can reduce synchronized extension and synthesis of the DNA strand and cause non-uniform sequence peak heights, affecting the read-length and possibly causing sequence errors. Studies have shown that the incorporation of dATP-α-S in T-homopolymeric regions results in uneven sequence signals and reduced sequence quality directly downstream of such homopolymers[12), 34)]. As mentioned earlier, during the development of Pyrosequencing technique, the natural dATP was replaced by dATP-α-S to increase the signal-to-noise ratio. Inefficient dATP-α-S incorporation by the exonuclease-deficient Klenow DNA polymerase caused the template extensions to go out of phase and cause asynchronous and ambiguous sequence peak signals that could be demonstrated in the pyrograms as uniformly reduced sequence signal peak-height. Therefore, efforts were made to compensate for this in the reading of those sections. By employing Sequenase[35)], an exonuclease deficient T7 DNA polymerase, the poly-T homopolymer strings reads were significantly improved and generated significantly more synchronized sequence and uniform signal peaks after homopolymeric T regions.

An important factor in Pyrosequencing is primer design for PCR and sequencing. Sequencing primers should be investigated for self-looping, primer-dimers (primer-primer hybridizations) and cross-hybridizations (when more than one sequencing primer is used). Single-stranded DNA binding protein (SSB)[36)] is very effective for primer and template complications in Pyrosequencing. Homopolymeric regions especially homopolymer T (complementary nucleotide A in nucleotide dispensation) are challenging due to Klenow polymerase and apyrase inefficiencies to incorporate fully and degrade the analogue dATP-α-S, resulting in non-synchronized extension and non-uniform peaks. Furthermore, it is highly recommended that sequencing primers to be designed before PCR primers in order to biotin-label the suitable DNA strand. This facilitates avoiding possible difficult homopolymers or regions that can interfere with efficient sequencing and base-callings.

5. Future of Pyrosequencing

DNA sequencing is one of the most central platforms for the study of biological systems. Future applications require more robust and efficient DNA sequencing techniques for sequence determination. Pyrosequencing is a relatively straightforward and user-friendly method possessing unique methodological characteristics and this technique is currently being used in multidisciplinary fields in academic, clinical and industrial settings. Pyrosequencing technology is a fairly novel DNA sequencing technique with room for improvements in chemistry and instrumentation. By increasing the read length to higher scores and by shortening the sequence reaction time per base calling, Pyrosequencing can take a more broad position in the DNA sequencing arena as the trend is directed to analysis of lesser amounts of specimens and large-scale settings, with higher throughput and lower cost. Miniaturization of this technique, and its integration with nucleic acid amplification and fully automated template preparation has made whole genome sequencing feasible and more significant advancements are expected for whole genome sequencing in near future.

References

1) Yager, T. D., Nickerson, D. A. and Hood, L. E. (1991) *Trends Biochem. Sci.* **16**, 454, 456, 458.
2) Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
3) Ronaghi, M., Uhlen, M. and Nyren, P. (1998) *Science* **281**, 363, 365.
4) Ahmadian, A. *et al.* (2000) Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* **280**, 103-110.
5) Nordstrom, T. *et al.* (2000) *Biotechnol. Appl. Biochem.* **31** (Pt 2), 107-112.
6) Gharizadeh, B., Kalantari, M., Garcia, C. A., Johansson, B. and Nyren, P. (2001) *Lab. Invest.* **81**, 673-679.
7) Nordstrom, T., Gharizadeh, B., Pourmand, N., Nyren, P. and Ronaghi, M. (2001) *Anal. Biochem.* **292**, 266-271.
8) Hyman, E. D. (1988) *Anal. Biochem.* **174**, 423-436.
9) Melamede, R. J. (1985) *U.S. Patent 4863849.*
10) Nyrén, P. (2001) *U.S. Patent 6,258,568B1.*
11) Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. and Nyren, P. (1996) *Anal. Biochem.* **242**, 84-89.
12) Gharizadeh, B., Nordstrom, T., Ahmadian, A., Ronaghi, M. and Nyren, P. (2002) *Anal. Biochem.* **301**, 82-90.
13) Milan, D. *et al.* (2000) *Science* **288**, 1248-1251.
14) Gharizadeh, B. *et al.* (2003) *Mol. Cell Probes* **17**, 203-210.
15) Grahn, N., Olofsson, M., Ellnebo-Svedlund, K., Monstein, H. J. and Jonasson, J. (2003) *FEMS Microbiol. Lett.* **219**, 87-91.
16) Jonasson, J., Olofsson, M. and Monstein, H. J. (2002) *Apmis.* **110**, 263-272.
17) Gharizadeh, B. *et al.* (2004) *Mycoses.* **47**, 29-33.
18) Trama, J. P., Mordechai, E. and Adelson, M. E. (2005) *J. Clin. Microbiol.* **43**, 906-908.
19) Trama, J. P., Mordechai, E. and Adelson, M. E. (2005) *Mol. Cell Probes* **19**, 145-152.
20) Adelson, M. E., Feola, M., Trama, J., Tilton, R. C. and Mordechai, E. (2005) *J. Clin. Virol.* **33**, 25-34.
21) Elahi, E. *et al.* (2003) *J. Virol. Methods* **109**, 171-176.
22) Gharizadeh, B. *et al.* (2003) *Nucleic Acids Res.* **31**, e146.
23) Gharizadeh, B. *et al.* (2005) *J. Mol. Diagn.* **7**, 198-205.
24) Ronaghi, M., Nygren, M., Lundeberg, J. and Nyren, P. (1999) *Anal. Biochem.* **267**, 65-71.
25) Ahmadian, A., Lundeberg, J., Nyren, P., Uhlen, M. and Ronaghi, M. (2000) *Biotechniques* **28**, 140-144, 146-147.
26) Garcia, C. A. *et al.* (2000) *Gene* **253**, 249-257.
27) Neve, B. *et al.* (2002) *Biotechniques* **32**, 1138-1142.
28) Rickert, A. M., Premstaller, A., Gebhardt, C. and Oefner, P. J. (2002) *Biotechniques* **32**, 592-593, 596-598, 600 *passim*.
29) Uhlmann, K., Brinckmann, A., Toliat, M. R., Ritter, H. and Nurnberg, P. (2002) *Electrophoresis* **23**, 4072-4079.
30) Gharizadeh, B. *et al.* (2003) *Electrophoresis* **24**, 1145-1151.
31) Gharizadeh, B. *et al.* (2006) *Mol. Cell Probes* **20**, 230-238.
32) Nourizad, N., Gharizadeh, B. and Nyren, P. (2003) *Electrophoresis* **24**, 1712-1715.
33) Margulies, M. *et al.* (2005) *Nature* **437**, 376-380.
34) Eriksson, J., Gharizadeh, B., Nourizad, N. and Nyren, P. (2004) *Nucleosides Nucleotides Nucleic Acids* **23**, 1583-1594.
35) Gharizadeh, B., Eriksson, J., Nourizad, N., Nordstrom, T. and Nyren, P. (2004) *Anal. Biochem.* **330**, 272-280.
36) Ronaghi, M. (2000) *Anal. Biochem.* **286**, 282-288.

Baback Gharizadeh

Baback Gharizadeh
Stanford Genome Technology Center, Stanford University, Research scientist
Address: 855 California Ave, Palo Alto, CA 94304, USA
E-mail: baback@stanford.edu
Mehran Ghaderi
Clinical pathology/cytology, Karolinska University Hospital, Research scientist
Address: Solna, SE-171 76 Stockholm, Sweden
Pål Nyrén
Department of Biotechnology, Royal Institute of Technology, Professor
Address: SE-106 91 Stockholm, Sweden

理論／
実験 技術