



Pyrosequencing Sheds Light on DNA Sequencing

Mostafa Ronaghi

Genome Res. 2001 11: 3-11

Access the most recent version at doi:[10.1101/gr.150601](https://doi.org/10.1101/gr.150601)

References

This article cites 31 articles, 10 of which can be accessed free at:
<http://genome.cshlp.org/content/11/1/3.full.html#ref-list-1>

Article cited in:

<http://genome.cshlp.org/content/11/1/3.full.html#related-urls>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>

Pyrosequencing Sheds Light on DNA Sequencing

Mostafa Ronaghi

Genome Technology Center, Stanford University, Palo Alto, California 94304, USA

DNA sequencing is one of the most important platforms for the study of biological systems today. Sequence determination is most commonly performed using dideoxy chain termination technology. Recently, pyrosequencing has emerged as a new sequencing methodology. This technique is a widely applicable, alternative technology for the detailed characterization of nucleic acids. Pyrosequencing has the potential advantages of accuracy, flexibility, parallel processing, and can be easily automated. Furthermore, the technique dispenses with the need for labeled primers, labeled nucleotides, and gel-electrophoresis. This article considers key features regarding different aspects of pyrosequencing technology, including the general principles, enzyme properties, sequencing modes, instrumentation, and potential applications.

The development of DNA sequence determination techniques with enhanced speed, sensitivity, and throughput are of utmost importance for the study of biological systems. Conventional DNA sequencing relies on the elegant principle of the dideoxy chain termination technique first described more than two decades ago (Sanger et al. 1977). This multi-step principle has gone through major improvements during the years to make it a robust technique that has been used for the sequencing of several different bacterial, archaeal, and eucaryotic genomes (<http://www.ncbi.nlm.nih.gov>, and <http://www.tigr.org>). However, this technique faces limitations in both throughput and cost for most future applications. Many research groups around the world have put effort into the development of alternative principles for DNA sequencing. Three methods that hold great promise are sequencing by hybridization (Bains and Smith 1988; Drmanac et al. 1989; Khrapko et al. 1989; Southern 1989), parallel signature sequencing based on ligation and cleavage (Brenner et al. 2000), and pyrosequencing (Ronaghi et al. 1996, 1998b). Pyrosequencing has been successful for both confirmatory sequencing and de novo sequencing. This technique has not been used for genome sequencing due to the limitation in the read length, but it has been employed for applications such as genotyping (Ahmadian et al. 2000a; Alderborn et al. 2000; Ekström et al. 2000; Nordström et al. 2000b), resequencing of diseased genes (Garcia et al. 2000), and sequence determination of difficult secondary DNA structure (Ronaghi et al. 1999). This article reviews the historical and technical aspects of the technique with regards to general principles, different strategies, application of the technique to different formats, and instrumentation. The performance of the technique in different applications is also discussed.

Pyrosequencing

Pyrosequencing is a DNA sequencing technique that is based on the detection of released pyrophosphate (PPi) during DNA synthesis. In a cascade of enzymatic reactions, visible light is generated that is proportional to the number of incorporated nucleotides (Fig. 1). The cascade starts with a nucleic acid polymerization reaction in which inorganic PPi is released as a result of nucleotide incorporation by polymerase. The released PPi is subsequently converted to ATP by ATP sulfurylase, which provides the energy to luciferase to oxidize luciferin and generate light. Because the added nucleotide is known, the sequence of the template can be determined. The nucleic acid molecule can be either RNA or DNA. However, because DNA polymerases show higher catalytic activity than RNA polymerases for limited nucleotide extension, efforts have been focused on the use of a primed DNA template for pyrosequencing. Standard pyrosequencing uses the Klenow fragment of *Escherichia coli* DNA Pol I, which is a relatively slow polymerase (Benkovic and Cameron 1995). The ATP sulfurylase used in pyrosequencing is a recombinant version from the yeast *Saccharomyces cerevisiae* (Karamohamed et al. 1999a) and the luciferase is from the American firefly *Photinus pyralis*. The overall reaction from polymerization to light detection takes place within 3–4 sec at room temperature. One pmol of DNA in a pyrosequencing reaction yields 6×10^{11} ATP molecules which, in turn, generate more than 6×10^9 photons at a wavelength of 560 nanometers. This amount of light is easily detected by a photodiode, photomultiplier tube, or a charge-coupled device camera (CCD) camera. There are two different pyrosequencing strategies that are currently available: solid-phase pyrosequencing (Ronaghi et al. 1996) and liquid-phase pyrosequencing (Ronaghi et al. 1998b). Solid-phase pyrosequencing (Fig. 2) utilizes immobilized DNA in the three-enzyme system described previously. In this system a washing step is performed to remove the excess substrate after each nucleotide addition. In

E-MAIL mostafa@stanford.edu; **FAX** (650) 812-1975.

Article and publication are at www.genome.org/cgi/doi/10.1101/gr.150601.

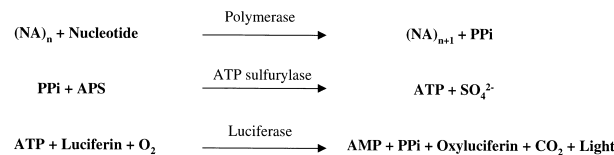


Figure 1 The general principle behind different pyrosequencing reaction systems. A polymerase catalyzes incorporation of nucleotide(s) into a nucleic acid chain. As a result of the incorporation, a pyrophosphate (PPi) molecule(s) is released and subsequently converted to ATP, by ATP sulfurylase. Light is produced in the luciferase reaction during which a luciferin molecule is oxidized.

liquid-phase pyrosequencing (Fig. 3) apyrase, a nucleotide-degrading enzyme from potato, is introduced to make a four-enzyme system. Addition of this enzyme has eliminated the need for solid support and intermediate washing thereby enabling the pyrosequencing reaction to be performed in a single tube. These formats are described in detail in this review.

History

The theory behind sequencing-by-synthesis was described in 1985 (Melamede 1985) and based on this principle, detection of pyrophosphate was used in DNA sequencing (Hyman 1988). Efforts were also put into the development of this principle for sequence determination using labeled nucleotides (Canard and Sarfati 1994; Cheesman 1994; Metzker et al. 1994; Rosenthal 1989; Tsien et al. 1991). However, Metzker et al. (1994) showed that the incorporation efficiency of labeled nucleotides is low, causing nonsynchronized extension, which made it difficult to sequence more than a few bases. Synchronized extension in sequencing-by-synthesis requires exonuclease-deficient (exo^-) DNA polymerase and unmodified nucleotides. We used coupled enzymatic reactions, which were used

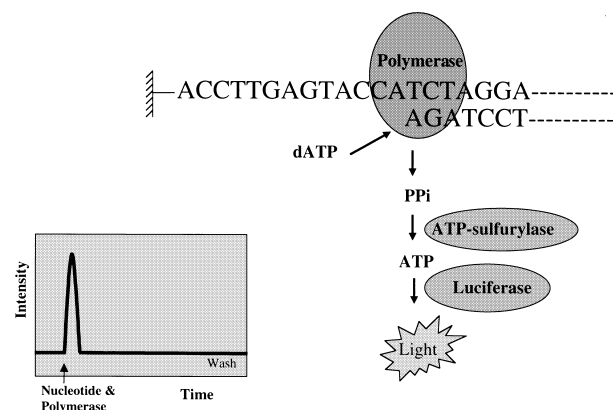


Figure 2 Schematic representation of the progress of the enzyme reaction in solid-phase pyrosequencing. The four different nucleotides are added stepwise to the immobilized primed DNA template and the incorporation event is followed using the enzyme ATP sulfurylase and luciferase. After each nucleotide addition, a washing step is performed to allow iterative addition.

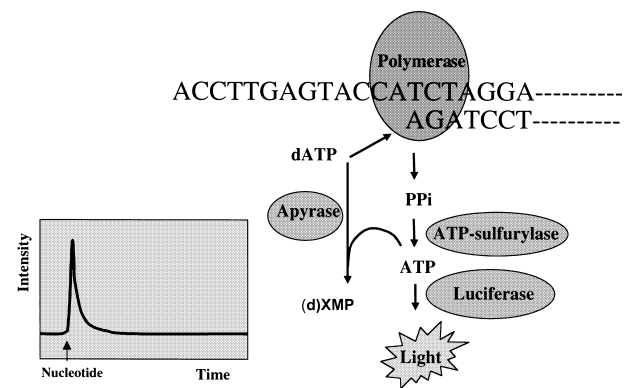


Figure 3 Schematic representation of the progress of the enzyme reaction in liquid-phase pyrosequencing. Primed DNA template and four enzymes involved in liquid-phase pyrosequencing are placed in a well of a microtiter plate. The four different nucleotides are added stepwise and incorporation is followed using the enzyme ATP sulfurylase and luciferase. The nucleotides are continuously degraded by nucleotide-degrading enzyme allowing addition of subsequent nucleotide. dXTP indicates one of the four nucleotides.

earlier to assay polymerase activity (Nyren 1987) to monitor stepwise DNA synthesis using exo^- polymerase and unlabeled nucleotides (pyrosequencing). However, false signals were always observed when dATP was added into the pyrosequencing solution (Ronaghi et al. 1996). The first major improvement was substitution of dATP α S for dATP in the polymerization reaction, which enabled the pyrosequencing reaction to be performed in homogeneous phase in real time (Ronaghi et al. 1996). It was later shown that the nonspecific signals were attributed to the fact that dATP is a substrate for luciferase. Conversely, dATP α S was found to be inert for luciferase, yet could be incorporated efficiently by all DNA polymerases tested (Ronaghi et al. 1996). This strategy was used successfully for sequencing of PCR-generated DNA material (Ronaghi et al. 1996).

The second improvement was the introduction of apyrase to the reaction to make a four-enzyme system (Ronaghi et al. 1998b). The addition of apyrase allowed nucleotides to be added sequentially without any intermediate washing step. This enzyme shows high catalytic activity and low amounts of this enzyme in the pyrosequencing reaction system efficiently degrade the unincorporated nucleoside triphosphates to nucleoside diphosphates and subsequently to nucleoside monophosphate. Apyrase is less inhibited by its products as compared to other nucleotide-degrading enzymes.

Most recently, the addition of ssDNA-binding protein to the pyrosequencing reaction system has simplified the optimization of different parameters in pyrosequencing. This protein has proven to be useful for long read sequencing and sequencing of difficult templates,

as well as providing flexibility in primer design (Ronaghi 2000).

Template Preparation for Pyrosequencing

Template preparation for pyrosequencing is straightforward. After generation of the template by PCR, the product should be purified prior to pyrosequencing. Unincorporated nucleotides and PCR primers in PCR reaction perturb the pyrosequencing reaction. The salt in the PCR reaction slightly inhibits the enzyme system and should be removed or diluted. Two strategies currently available for generation of a primed DNA template for pyrosequencing are described below.

Solid-Phase Template Preparation

Streptavidin-coated magnetic beads have been used to prepare primed DNA template for pyrosequencing. This technology enables biotinylated PCR product to be captured onto magnetic beads. After sedimentation, the remaining components of the PCR reaction can be removed by washing to obtain pure double-stranded DNA followed by alkali denaturation to yield ssDNA. Both the immobilized biotinylated and nonbiotinylated strands in solution can be used as pyrosequencing templates (Ronaghi et al. 1998a, 1999). This template preparation system has given high-quality sequence data with low background signals.

Enzymatic Template Preparation

Recently, enzymatic template preparation was developed for sequencing on double-stranded DNA template (Nordstrom et al. 2000a,b). This template preparation method employs a nucleotide-degrading enzyme and exonuclease I. The enzymes are added to the PCR product and the mixture is incubated at room temperature or 35°C. During this step, the nucleotide-degrading enzyme removes the nucleotides and exonuclease I degrades the PCR primers remaining from the amplification step. The sequencing primer is dispensed into the treated mixture and the temperature of the solution is increased to heat-inactivate the enzymes. Template/primer complexes are formed by rapid cooling of the solution. Two different enzyme systems can be used. The use of alkaline phosphatase from shrimp or calf intestine together with exonuclease I allows the template to be prepared within 20 min, whereas a combination of a low amount of apyrase, inorganic PPi, and exonuclease I enables the template to be prepared in three min. High quality pyrosequencing data has been obtained by enzymatic template preparation using a prototype pyrosequencing system that employs a very sensitive light detector. However, this template preparation method needs to be further optimized for use with the standard system that uses microtiter plates, because the dilution that is required to compensate for incompatible buffer systems results in low amounts of primed DNA template. Improve-

ments may be obtained by running PCR in a compatible buffer with pyrosequencing reaction or by using a more sensitive CCD camera in the pyrosequencing machine.

Pyrosequencing Enzyme Systems

Pyrosequencing takes advantage of the cooperativity of several enzymes to monitor DNA synthesis. Parameters such as stability, fidelity, specificity, sensitivity, K_M , and k_{cat} (Table 1) are of utmost importance for the optimal performance of the enzymes used in the reaction (Ronaghi 1998). The kinetics of the enzymes can be studied in real time by following the pyrosequencing signals (a pyrogram). The slope of the ascending curve in a pyrogram (Figs. 4 and 6) is determined mainly by the activities of polymerase and ATP sulfurylase; the height of the signal is determined by the activity of luciferase, and the slope of the descending curve by the efficiency of nucleotide removal. In the solid-phase system using microfluidics, which employs the three-enzyme system, the descending curve is determined by the washing efficiency. In the four-enzyme system of liquid-phase pyrosequencing, the accumulation of inhibitory substances decreases the efficiency of luciferase and apyrase. In both systems, the activity of ATP sulfurylase is relatively constant during the sequencing reaction. In pyrosequencing, the most critical reactions are DNA polymerization and nucleotide removal by either washing or enzymatic degradation. Nucleotide removal (descending curve) competes with the polymerization reaction (ascending curve). Therefore, slight changes in the kinetics of these reactions directly influence the performance of the sequencing reaction.

Polymerization Reaction

An excess amount of DNA polymerase relative to DNA template in the pyrosequencing reaction ensures that the primed DNA template is bound efficiently by polymerase and that, at the time of nucleotide addition, polymerization takes place immediately. To obtain rapid polymerization, the nucleotide concentration must be above the K_M of the DNA polymerase (Table 1).

Table 1. Kinetic Data of Enzymes Involved in Pyrosequencing

Enzyme	K_M (μ M)	k_{cat} (S^{-1})
Klenow Polymerase ^a	0.18 (dTTP)	0.92
ATP sulfurylase ^b	0.56 (APS) 7.0 (PPi)	38
Firefly luciferase ^c	20 (ATP)	0.015
Apyrase ^d	120 (ATP) 260 (ADP)	500 (ATP)

^aVan Draanen et al. 1992.

^bNyren and Lundin 1985.

^cDeLuca and McElroy 1984.

^dTraverso-Cori et al. 1965.

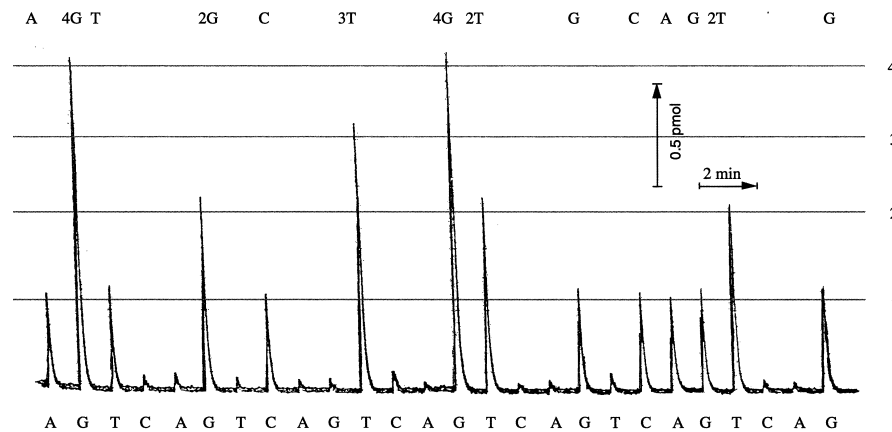


Figure 4 Pyrogram of the raw data obtained from liquid-phase pyrosequencing. Proportional signals are obtained for one, two, three, and four base incorporations. Nucleotide addition, according to the order of nucleotides, is indicated below the pyrogram and the obtained sequence is indicated above the pyrogram.

Conversely, if the concentration of the nucleotides is too high, lower fidelity of the polymerase is observed (Eckert and Kunkel 1990; Cline et al. 1996), even though the K_M for misincorporation is much higher than that of correct incorporation (Gillin and Nossal 1976; Topal et al. 1980; Capson et al. 1992). A high fidelity can be achieved by using polymerases with inherent exonuclease activity, however, this has the disadvantage that primer degradation can occur causing out of phase signals. Although the exonuclease activity of Klenow polymerase is relatively low, it has been found that the 3' end of the primer was degraded during long incubations in the absence of nucleotides. Even without exonuclease activity, an induced-fit binding mechanism in the polymerization step (Wong et al. 1991) provides a very high selectivity for the correct nucleotide, with a fidelity of 10^5 – 10^6 when the nucleotide are added slightly above the K_M . In pyrosequencing, exo-polymerases, such as exo-Klenow or Sequenase, catalyze the incorporation of a nucleotide only in the presence of a complementary nucleotide, confirming the high fidelity of these enzymes even in the absence of proof-reading exonuclease activity. The K_M and k_{cat} for one-base incorporation is lower than that for the incorporation of several bases for most polymerases (Van Draanen et al. 1992). However, the K_M values for nucleotides are much lower for DNA polymerases than for apyrase (Table 1). Therefore, an increased fidelity in the system can be obtained as the nucleotide concentration necessary for efficient polymerization is relatively low, and apyrase degrades nucleotides to a concentration far below the K_M of the polymerase in less than five sec.

Detection Enzymes

On successful polymerization, a proportional amount of PPI is released. ATP sulfurylase converts PPI to ATP in

~1.5 sec and the generation of light by luciferase takes place in <0.2 sec. In the four-enzyme system accumulation of AMP and dAMP α S inhibits the luciferase activity. Kinetics of the enzymes in the detection reaction can be followed by the addition of a known amount of PPI to the pyrosequencing enzyme system.

Nucleotide Removal

To allow iterative addition, nucleotides must be removed from the pyrosequencing reaction. In the four-enzyme system, nucleotides are removed enzymatically. The nucleotide-degrading enzyme must possess

the following properties: First, the enzyme must hydrolyze all deoxynucleoside triphosphates at approximately the same rate; second, it must hydrolyze ATP to prevent the accumulation of ATP between cycles; third, the time for nucleotide degradation by the nucleotide-degrading enzyme must be slower than nucleotide incorporation by the polymerase. It is also important that the yield of primer-directed incorporation is as close to 100% as possible before the nucleotide-degrading enzyme has degraded the nucleotide to a concentration below the K_M of the polymerase (Table 1). Pyrograms obtained from liquid-phase pyrosequencing (Fig. 4) show that apyrase fulfils the criteria described above. A constant signal intensity for each base incorporation is obtained during the course of a reaction determining high efficiency of this system (Fig. 4). In solid-phase pyrosequencing (three-enzyme

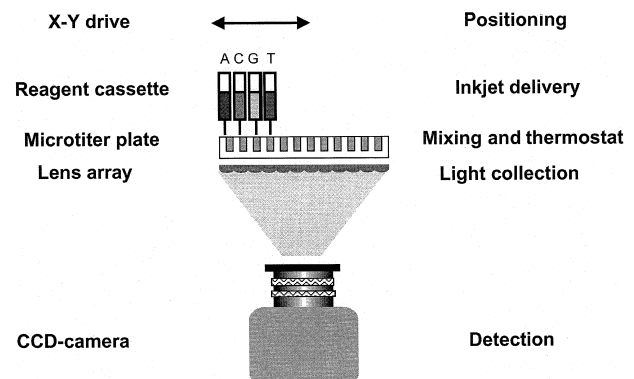


Figure 5 Schematic drawing of the automated system for liquid-phase pyrosequencing. Four dispensers move on an X-Y robotics arm over the microtiter plate and add four different nucleotides, according to the prespecified order. The microtiter plate is agitated continuously to mix the added nucleotide. Generated light is directed to the CCD camera using a lens array located exactly below the microtiter plate.

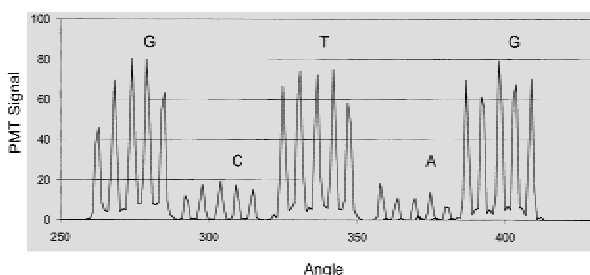


Figure 6 Pyrogram obtained from five different chambers using microfluidics in a spinning CD-format device developed at Gyros AB (Sweden). All chambers contained the same DNA template. The addition order of nucleotides was GCTA. Nucleotides G, T, and G were correctly incorporated. Background signals for nucleotides C and A were due to pyrophosphate (PPi) contamination in the pyrosequencing reaction mixture.

system), washing in a controlled manner should show the same advantages offered by apyrase in the four-enzyme system. An additional advantage of the three-enzyme system is that no accumulation of inhibitory substances will be observed, because washing is performed between each nucleotide addition.

Extending the Read Length of Pyrosequencing

For many applications such as genome sequencing and gene sequencing, a long read is desirable. Several criteria must be met to obtain a long read length in pyrosequencing: (1) The enzyme system must be stable; (2) there must be a low misincorporation; and (3) nucleotide extension must be synchronized. The enzyme system has been shown to be stable in its buffer system during the sequencing reaction as relatively constant signal intensity is obtained for each individual nucleotide (Fig. 4). In the four-enzyme system, removal of inhibitory substances from the reaction and minimizing the dilution effect gives rise to 200 nucleotide reads (Ronaghi et al. 2000). The use of unlabeled nucleotides, addition of nucleotide in a concentration slightly above the K_M , and rapid removal of nucleotides from the solution increases the fidelity of DNA synthesis. Although a relatively high signal-to-noise ratio is obtained in pyrosequencing, misincorporation may play an important role in limiting longer reads. Possible misincorporation terminates the primer strands, which results in decreased signal intensity in the course of a reaction. Nonsynchronized extensions are either a result of minus frame shift (when some of the primer strands get one, or a few, nucleotides behind the other synchronized primer strands during extension) or plus frame shift (when some of the primer strands get one, or a few, nucleotides ahead of other synchronized primer strands during extension). Using exo^- DNA polymerase reduces the minus frame shift. Insufficient exposure of nucleotides to polymerase can cause minus frame shift, which is sometimes observed in long homopolymeric regions. In the four-enzyme

system of pyrosequencing, apyrase degrades the nucleotides below the K_M for polymerase, not allowing enough time for the polymerase to complete the polymerization of these regions. The use of lower amounts of apyrase, or a second addition of the same nucleotide, solves this problem. Improvements can also be obtained by the use of ssDNA-binding protein in the pyrosequencing reaction solution (Ronaghi 2000). In a microfluidic format, complete incorporation can be controlled easily by a delay in washing. Plus frame shift is mainly a problem for the four-enzyme system of pyrosequencing and normally is caused by enzyme contaminants or inefficient nucleotide degradation. A contaminating enzyme such as nucleoside diphosphate kinase, which normally is found in commercially available ATP sulfurylase and apyrase, converts the nucleoside diphosphate to nucleoside triphosphate (Karamohamed et al. 1999b), a substrate for polymerase. Another parameter causing plus frame shift is inefficient degradation of nucleotides by apyrase and usually is seen in later cycles of pyrosequencing due to the accumulation of inhibitory substances. Further works are underway to remove the inhibitory substances from the reaction system either by purification or enzymatic degradation. Another factor reducing the efficiency of the four-enzyme system is the dilution effect. In the four-enzyme system, the nucleotides are iteratively added to the pyrosequencing solution thereby increasing the reaction volume at each step. Although the volume of nucleotides added is as little as 200 nanoliter/min, dilution can be seen in long-read sequencing. Dilution lowers the enzyme concentrations thereby decreasing their efficiency. Possible improvements include reducing the volume of nucleotide delivery or running the reaction at higher temperatures to increase evaporation.

Challenges for Pyrosequencing Technology

An inherent problem with the described method is de novo sequencing of polymorphic regions in heterozygous DNA material. In most cases, it will be possible to detect the polymorphism. If the polymorphism is a substitution, it will be possible to obtain a synchronized extension after the substituted nucleotide. If the polymorphism is a deletion or insertion of the same kind as the adjacent nucleotide on the DNA template, the sequence after the polymorphism will be synchronized. However, if the polymorphism is a deletion or insertion of another type, the sequencing reaction can become out of phase, making the interpretation of the subsequent sequence difficult. If the polymorphism is known, it is always possible to use programmed nucleotide delivery to keep the extension of different alleles synchronized after the polymorphic region. It is also possible to use a bidirectional approach (Ronaghi et al. 1999) whereby the complementary strand is se-

quenced to decipher the sequence flanking the polymorphism.

Another inherent problem is the difficulty in determining the number of incorporated nucleotides in homopolymeric regions, due to the nonlinear light response following incorporation of more than 5–6 identical nucleotides. The polymerization efficiency over homopolymeric regions has been investigated and the results indicate that it is possible to incorporate ≤ 10 identical adjacent nucleotides in the presence of apyrase (Ronaghi 2000). However, to elucidate the correct number of incorporated nucleotides, it may be necessary to use specific software algorithms that integrate the signals. For resequencing, it is possible to add the nucleotide twice for a homopolymeric region to ensure complete polymerization.

Applications of Pyrosequencing

Pyrosequencing has opened up new possibilities for performing sequence-based DNA analysis. The availability of an automated system for liquid-phase pyrosequencing (PSQ 96 system, <http://www.pyrosequencing.com>) has allowed the technique to be adapted for high-throughput analyses. This section describes some of the potential applications of pyrosequencing.

Genotyping of Single-Nucleotide Polymorphisms

For analysis of single-nucleotide polymorphisms (SNPs) by pyrosequencing, the 3'-end of a primer is designed to hybridize one or a few bases before the polymorphic position. In a single tube, all the different variations can be determined as the region is sequenced. A striking feature of pyrogram readouts for SNP analysis is the clear distinction between the various genotypes; each allele combination (homozygous or heterozygous) will give a specific pattern compared with the two other variants (Ahmadian et al. 2000a; Alderborn et al. 2000; Ekstrom et al. 2000; Nordstrom et al. 2000b). This feature makes typing extremely accurate and easy. Relative standard deviation values for the ratio between key peaks of the respective SNPs and reference counterparts are ≤ 0.1 (Alderborn et al. 2000). Simple manual comparison of predicted SNP patterns and the raw data obtained from the PSQ 96 system can score an SNP, especially as no editing is needed. Because specific patterns can be readily achieved for the individual SNPs, it will also be possible to automatically score the allelic status by pattern recognition software. In a study based on results from three different laboratories, 26 different SNPs and >1600 DNA samples were analyzed. The algorithm classified the data from 94% of the samples as good or medium quality and 99.4% of these were automatically assigned the expected genotypes (B. Ekström, pers. comm.). The major reason for low quality data was insufficient signal/noise typically caused by low efficiency in PCR ampli-

fication. As pyrosequencing signals are very quantitative, it is possible to use this strategy for the studies of allelic frequency in large population. This system allows >5000 samples to be analyzed in 8 h. Furthermore, pyrosequencing enables determination of the phase of SNPs when they are in the vicinity of each other allowing the detection of haplotypes (Ahmadian et al. 2000b).

Microbial Typing

DNA markers used for typing normally contain both conserved and variable regions. A DNA primer complementary to the conserved or semiconserved region is usually employed to sequence the variable region. In bacteria, 16S rRNA gene is commonly used to identify different species and strains. By analyzing a sequence between 20–100 nucleotides on 16S rRNA gene, it is possible to taxonomically group different bacteria and, in many cases it is possible to get information about the strains. Pyrosequencing is now being applied for rapid typing of large number of bacteria, yeasts, and viruses (B. Gharizadeh, pers. comm.).

Resequencing

Pyrosequencing is currently the fastest method for sequencing a PCR product. Because pyrosequencing generates an accurate quantification of the mutated nucleotides, the resequencing of PCR-amplified disease genes for mutation scanning will be one of the more interesting applications. Using this technique for resequencing results in longer read length than de novo sequencing because nucleotide delivery can be specified according to the order of the sequence. Programmed dispensing generates a signal for each addition in a pyrogram, therefore variation in the pattern indicates the appearance of a mutation. This strategy has been used for resequencing of the p53 tumor suppressor gene where mutations were successfully determined and quantified (Garcia et al. 2000).

Tag Sequencing

The sequence order of nucleotides determines the nature of the DNA. Theoretically, eight or nine nucleotides in a row should define a unique sequence for every gene in the human genome. However, it has been found that to uniquely identify a gene from a complex organism such as human, a longer sequence of DNA is needed. In a pilot study, it was found that 98% of genes in a human cDNA library could be uniquely identified by sequencing a length of 30 nucleotides. Pyrosequencing was used to sequence this length for gene identification from a human cDNA library and the results were in complete agreement with longer sequence data obtained by Sanger DNA sequencing. Pyrosequencing offers high-throughput analysis of cDNA libraries because 96 samples can be analyzed in less than one hour. Like Sanger DNA se-

quencing, pyrosequencing also has the advantage of library screening, as the original cDNA clone is directly available for further analysis.

Analysis of Difficult Secondary Structures

Hairpin structures are common features in genomic material and have been proposed to have regulatory functions in gene transcription and replication. However, analyzing these sequences by conventional DNA sequencing usually gives rise to DNA sequence ambiguities seen as "run-off" or compressions. These problems have been associated with gel electrophoresis. Pyrosequencing was successfully applied to decipher the sequence of such regions (Ronaghi et al. 1999). Klenow DNA polymerase was used in these studies in which relatively high strand-displacement activity in reading through these structures was shown.

Instrumentation

Automation Based on Microtiter Plate Format

An automated version of a pyrosequencing machine was recently developed (<http://www.pyrosequencing.com>). The automated version uses a disposable inkjet cartridge for precise delivery of small volume (200 nL) of six different reagents into a temperature-controlled microtiter plate (Fig. 5). The microtiter plate is under continuous agitation to increase the rate of the reactions. A lens array is used to efficiently focus the generated luminescence from each individual well of the microtiter plate onto the chip of a CCD camera. Nucleotides are dispensed into alternating wells with a delay to minimize the crosstalk of generated light between different wells. A cooled CCD camera images the plate every second to follow the exact process of the pyrosequencing reaction. Data acquisition modules and an interface for PC connection are used in this instrument. Software running under Windows NT enables individual control of the dispensing order for each well. Prior to pyrosequencing, the reagents and each of the four nucleotides are loaded into the inkjet cartridge that is mounted in the instrument. A microtiter plate containing primed DNA template is placed into the pyrosequencing machine, and after the enzymes and substrate have been delivered by the inkjet, nucleotides are added to the solution according to the specified order. The signals in a pyrogram (Fig. 4) show high quality sequence data with high signal-to-noise ratio with the height of the peaks proportional to the number of incorporated nucleotides. A high-throughput version of this machine is also under development, which will allow the analysis of $\leq 50,000$ SNPs per day (B. Ekström, pers. comm.).

Microfluidics Using Solid-Phase Pyrosequencing

Running pyrosequencing on solid-phase in a microfluidic system offers several advantages compared with

other described formats: (1) Sequencing can be performed faster because the cycling time for each nucleotide addition can be reduced; (2) there is no accumulation of inhibitory substances because washing is performed after each cycle; (3) it is possible to use lower amounts of enzymes to reduce the cost, and (4) integration of PCR amplification, template preparation, and pyrosequencing analysis in a single flow system can be envisaged. Using this format, DNA templates are immobilized on a solid support that enables iterative washing. Eventually it may be possible to immobilize the detection enzymes (ATP sulfurylase and luciferase) onto a solid support to further reduce enzyme consumption. Pyrosequencing in a microfluidic format has the potential to be used for very long reads because no accumulation of inhibitory substances will be obtained. Different microfluidic formats are currently being tested for pyrosequencing analysis. The pyrogram in Figure 6 demonstrates promising sequencing data with a relatively high signal-to-noise ratio that was obtained in a centrifugal force-driven compact disc microfluidic device (Eckersten et al. 2000).

Array Pyrosequencing

Developments in microarray technology have opened new possibilities for sequence-based analysis. The major advantages of these formats are low cost and high throughput. Pyrosequencing can be applied on both ordered and random arrays. In the latter array format, pyrosequencing data provides high amounts of information to reveal the sequence of the DNA template to be analyzed, eliminating the decoding step. A similar strategy can be applied to tag sequencing of a cDNA library immobilized on a solid surface using a common sequencing primer. A system has been built that employs a nucleotide delivery module, a DNA array, and a CCD camera (M. Ronaghi N. Pourmand, M. Jain, T. Willis, and R.W. Davis, in prep.). A piezoelectric ultrasonic sprayer was recently developed to enable homogenous delivery of nucleotide onto a microarray. A single sprayer is used to deliver all four different nucleotides, with washing of the nozzle between each delivery. Data from single-base extension on an oligonucleotide template attached to a glass slide has been obtained, showing the feasibility of the pyrosequencing enzymatic reactions in this format. It should be noted that in this chemiluminescence assay the energy available for detection is proportional to the amount of template in the reaction, whereas in fluorescence assay the energy available for detection can be increased through the use of more powerful lasers. Consequently, sensitive detection systems must be employed to allow detection of miniature pyrosequencing reaction. Imaging a pyrosequencing reaction onto currently available CCD technology, we believe that the smallest detectable reaction should contain ≥ 5000

template molecules. Further optimization needs to be performed in terms of diffusion and variability of the amount of available primed DNA templates before application of such a format for reliable high throughput DNA sequencing.

Software for Pyrogram Analysis

Specialized software for SNP analysis of pyrograms obtained from liquid-phase pyrosequencing has been developed (<http://www.pyrosequencing.com>). This software enables analyses of selected wells in a microtiter plate and automatically performs genotyping as well as quality assessment of the raw data utilizing a novel SNP genotyping algorithm. Based on pattern recognition, this algorithm automatically scores the genotype and calculates a quality value for each SNP scored. The assignment of quality values is based on a number of different parameters, including difference in match between the best and next best choice of genotypes, agreement between expected and obtained sequence around the SNP, signal-to-noise ratios, variance in peak heights around the SNP, and peak width. This software has also been used for other applications such as EST sequencing, microbial typing, and confirmatory sequencing, however, until now the base-calling has been performed manually. Specialized software for pyrosequencing of longer reads is currently under development with automatic base-calling (B. Ekström, pers. comm.).

Future Trends

Genome sequencing provides tremendous amounts of information that can be used in several different areas of biology. Comparative sequencing will dominate DNA sequencing to identify variations across those genomes that have been sequenced. Technologies with high accuracy for the identification of these variations in genome-wide scanning will have great value. Pyrosequencing has shown excellent accuracy for analysis of polymorphic DNA fragments. This technology has also been used for quantification of allelic frequency in populations. While the variations are characterized, correlation of variation to phenotype can be performed. Pyrosequencing will have a large impact in that area because a large number of samples can be pooled in one pyrosequencing reaction. A high throughput version of this technology can potentially be used for resequencing of genomes.

Pyrosequencing technology is relatively new and there is much room for developments in both chemistry and in instrumentation. The technology is already time- and cost-competitive (the cost is currently 69 cents per sample using standard pyrosequencing; <http://www.pyrosequencing.com>) when compared with the existing sequencing methods. Work is underway to further improve the chemistry, to measure the sequencing efficiency at elevated temperatures, and to

run the reaction in miniaturized formats. The advantage of pyrosequencing in miniaturized formats may lie in the ease with which large numbers of high-density arrays can be manufactured and the future integration of sample preparation with these devices. Success in miniaturization of this technique into high density microtiter plates, microarrays, or microfluidics will reduce the cost and increase the throughput by one to two orders of magnitude, a crucial step for large scale genetic testing.

ACKNOWLEDGMENTS

The author is supported by an NIH grant. I thank Ronald Davis for valuable discussions, Guri Glaever, Joakim Lundberg, Paul Hardenbol, Thomas Willis, Pål Nyrén, and Björn Ekström for valuable comments on this manuscript. I also thank Baback Gharizadeh, Afshin Ahmadian, Nader Pourmand, and Nigel Tooke for sharing their results on pyrosequencing.

REFERENCES

- Ahmadian, A., Gharizadeh, B., Gustafsson, A.C., Sterky, F., Nyren, P., Uhlen, M., and Lundberg, J. 2000a. Single-nucleotide polymorphism analysis by Pyrosequencing. *Anal. Biochem.* **280**: 103–110.
- Ahmadian, A., Lundberg, J., Nyren, P., Uhlen, M., and Ronaghi, M. 2000b. Analysis of the p53 tumor suppressor gene by pyrosequencing. *BioTechniques* **28**: 140–144.
- Alderborn, A., Kristofferson, A., and Hammerling, U. 2000. Determination of single nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Res.* **10**: 1249–1258.
- Bains, W. and Smith, G.C. 1988. A novel method for nucleic acid sequence determination. *J. Theoret. Biol.* **135**: 303–307.
- Benkovic, S.J. and Cameron, C.E. 1995. Kinetic analysis of nucleotide incorporation and misincorporation by Klenow fragment of *Escherichia coli* DNA polymerase I. *Methods Enzymol.* **262**: 257–269.
- Brenner, S., Williams, S.R., Vermaas, E.H., Storck, T., Moon, K., McCollum, C., Mao, J.L., Luo, S., Kirchner, J.J., Eletr, S., et al. 2000. In vitro cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs. *Proc. Natl. Acad. Sci.* **97**: 1665–1670.
- Canard, B. and Sarfati, R.S. 1994. DNA polymerase fluorescent substrates with reversible 3'-tags. *Gene* **148**: 1–6.
- Capson, T.L., Peliska, J.A., Kaboord, B.F., Frey, M.W., Lively, C., Dahlberg, M., and Benkovic, S.J. 1992. Kinetic characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4. *Biochemistry* **31**: 10984–10994.
- Cheesman, P.C. 1994. Method for sequencing polynucleotides. US Patent no. 5302509.
- Cline, J., Braman, J.C., and Hogrefe, H.H. 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* **24**: 3546–3551.
- DeLuca, M. and McElroy, W.D. 1984. Two kinetically distinguishable ATP sites in firefly luciferase. *Biochem. Biophys. Res. Commun.* **123**: 764–770.
- Drmanac, R., Labat, I., Brukner, I., and Crkvenjakov, R. 1989. Sequencing of megabase plus DNA by hybridization: Theory of the method. *Genomics* **4**: 114–128.
- Eckert, A., Örfors, A.E., Ellström, C., Erickson, A., Löfman, E., Eriksson, A., Eriksson, S., Jorsback, A., Tooke, N., Derand, H., et al. 2000. High-throughput SNP scoring in a disposable microfabricated CD device. In *Proceedings of the Micro Total Analysis Systems*. Kluwer Academic Publishers.
- Eckert, K.A. and Kunkel, T.A. 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res.* **18**: 3739–3744.

- Ekstrom, B., Alderborn, A., and Hammerling, U. 2000. Pyrosequencing for SNPs. *Progress in biomedical optics* **1**: 134–139.
- Garcia, A.C., Ahamdian, A., Gharizadeh, B., Lundeberg, J., Ronaghi, M., and Nyren, P. 2000. Mutation detection by Pyrosequencing: Sequencing of exons 5 to 8 of the p53 tumour suppressor gene. *Gene* **253**: 249–257.
- Gillin, F.D. and Nossal, N.G. 1976. Control of mutation frequency by bacteriophage T4 DNA polymerase. II. Accuracy of nucleotide selection by the L88 mutator, CB120 antimutator, and wild type phage T4 DNA polymerases. *J. Biol. Chem.* **251**: 5225–5232.
- Hyman, E.D. 1988. A new method of sequencing DNA. *Anal. Biochem.* **174**: 423–436.
- Karamohamed, S., Nilsson, J., Nourizad, K., Ronaghi, M., Pettersson, B., and Nyren, P. 1999a. Production, purification, and luminometric analysis of recombinant *Saccharomyces cerevisiae* MET3 adenosine triphosphate sulfurylase expressed in *Escherichia coli*. *Prot. Exp. Purif.* **15**: 381–388.
- Karamohamed, S., Nordstrom, T., and Nyren, P. 1999b. Real-time bioluminometric method for detection of nucleoside diphosphate kinase activity. *BioTechniques* **26**: 728–734.
- Khrapko, K.R., Lysov Yu, P., Khorlyn, A.A., Shick, V.V., Florentiev, V.L., and Mirzabekov, A.D. 1989. An oligonucleotide hybridization approach to DNA sequencing. *FEBS Lett.* **256**: 118–122.
- Melamede, R.J. 1985. Automatable process for sequencing nucleotide. US Patent no. US4863849.
- Metzker, M.L., Raghavachari, R., Richards, S., Jacutin, S.E., Civitello, A., Burgess, K., and Gibbs, R.A. 1994. Termination of DNA synthesis by novel 3'-modified-deoxyribonucleoside 5'-triphosphates. *Nucleic Acids Res.* **22**: 4259–4267.
- Nordstrom, T., Nourizad, K., Ronaghi, M., and Nyren, P. 2000a. Methods enabling Pyrosequencing on double-stranded DNA. *Anal. Biochem.* **282**: 186–193.
- Nordstrom, T., Ronaghi, M., Forsberg, L., de Faire, U., Morgenstern, R., and Nyren, P. 2000b. Direct analysis of single-nucleotide polymorphism on double-stranded DNA by pyrosequencing. *Biotechnol. Appl. Biochem.* **31**: 107–112.
- Nyren, P. 1987. Enzymatic method for continuous monitoring of DNA polymerase activity. *Anal. Biochem.* **167**: 235–238.
- Nyren, P. and Lundin, A. 1985. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Anal. Biochem.* **151**: 504–509.
- Ronaghi, M. 1998. 'Pyrosequencing: A tool for sequence-based DNA analysis.' Doctoral thesis, The Royal Institute of Technology, Stockholm, Sweden.
- . 2000. Improved performance of Pyrosequencing using single-stranded DNA-binding protein. *Anal. Biochem.* **286**: 282–288.
- Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M., and Nyren, P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**: 84–89.
- Ronaghi, M., Pettersson, B., Uhlen, M., and Nyren, P. 1998a. PCR-introduced loop structure as primer in DNA sequencing. *BioTechniques* **25**: 876–884.
- Ronaghi, M., Uhlen, M., and Nyren, P. 1998b. A sequencing method based on real-time pyrophosphate. *Science* **281**: 363–365.
- Ronaghi, M., Nygren, M., Lundeberg, J., and Nyren, P. 1999. Analyses of secondary structures in DNA by pyrosequencing. *Anal. Biochem.* **267**: 65–71.
- Ronaghi, M., Pourmand, N., Jain, M., Willis, T., and Davis, R. 2000. Pyrosequencing for genome resequencing. In *12th International Genome Sequencing and Analysis Conference*, Miami, FL.
- Rosenthal, A. 1989. Process for solid phase-sequencing of nucleic acid. US Patent no. US1985000761107.
- Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463–5467.
- Southern, E.M. 1989. Analysing polynucleotide sequences. US Patent no. WO/10977.
- Topal, M.D., DiGiuseppi, S.R., and Sinha, N.K. 1980. Molecular basis for substitution mutations. Effect of primer terminal and template residues on nucleotide selection by phage T4 DNA polymerase in vitro. *J. Biol. Chem.* **255**: 11717–11724.
- Traverso-Cori, A., Chaimovich, H., and Cori, O. 1965. Kinetic studies and properties of potato apyrase. *Arch. Biochem. Biophys.* **109**: 173–181.
- Tsien, R.Y., Ross, P., Fahnestock, M., and Johnston, A.J. 1991. Method for DNA sequencing. US Patent no. PCT WO 91/06678.
- Van Draanen, N.A., Tucker, S.C., Boyd, F.L., Trotter, B.W., and Reardon, J.E. 1992. Beta-L-thymidine 5'-triphosphate analogs as DNA polymerase substrates. *J. Biol. Chem.* **267**: 25019–25024.
- Wong, L., Patel, S.S., and Johnson, K.A. 1991. An induced-fit kinetic mechanism for DNA replication fidelity: Direct measurement by single-turnover kinetics. *Biochemistry* **30**: 526–537.