Published in final edited form as:

Anal Biochem. 2007 April 15; 363(2): 275–287.

Analysis of Read-Length Limiting Factors in Pyrosequencing Chemistry

Foad Mashayekhi and Mostafa Ronaghi*

Stanford Genome Technology Center, Stanford University, Palo Alto, CA, USA

Abstract

Pyrosequencing is a bioluminometric DNA sequencing technique that measures the release of pyrophosphate during DNA synthesis. The amount of pyrophosphate is proportionally converted into visible light by a cascade of enzymatic reactions. Pyrosequencing has thus far been used for generating short sequence reads (1-100 nucleotides), as certain factors limit the system's ability to accurately perform longer reads. In this study, we have characterized the main read-length limiting factors in both three-enzyme and four-enzyme Pyrosequencing systems. A new simulation model was developed to simulate the read-length of both systems, based on the inhibitory factors in the chemical equations governing each enzymatic cascade. Our results indicate that non-synchronized extension limits the obtained read-length; however, to a different extent for each system. In four-enzyme system, non-synchronized extension due mainly to a decrease in apyrase's efficiency in degrading excess nucleotides proves to be the main limiting factor of read-length. Replacing apyrase with a washing step for removal of excess nucleotide proves essential in improving the read-length of Pyrosequencing. The main limiting factor of the three-enzyme system is shown to be loss of DNA fragments during the washing step. If this loss is minimized to 0.1% per washing cycle, the read-length of Pyrosequencing would be well beyond 300 bases.

Keywords

Pyrosequencing; sequencing-by-synthesis; enzyme kinetics; read-length; DNA sequencing; enzyme simulation

INTRODUCTION

Developing vastly improved DNA sequencing techniques would direct the revolution initiated by the Human Genome Project toward low-cost, high-throughput whole-genome sequencing, personalized medicine, and other related projects such as ecological studies. The Human Genome Project was made possible by a reduction in DNA sequencing cost by three orders of magnitude. It is believed that further cost reduction by two to three orders of magnitude will be necessary to enter a new era of DNA sequencing applications. Pyrosequencing (1,2) has emerged as one of the major non-gel-based DNA sequencing methods for short reads and whole-genome sequencing. Also, since its approval as a standard technique by NCBI, Pyrosequencing is likely to gain a more dominant position in the sequencing arena.

CORRESPONDING AUTHOR: Mostafa Ronaghi, 855 California Ave, Palo Alto, CA 94304, Email: mostafa@stanford.edu, Phone: (650) 812-1971, Fax: (650) 812-1975.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Pyrosequencing is a sequencing-by-synthesis method. The four different nucleotides, dATP, dCTP, dGTP, and dTTP, are added one by one. Pyrosequencing relies on the real-time detection of the inorganic pyrophosphate (1) released upon incorporation of complementary nucleotides during DNA synthesis. Generated PPI is immediately converted to adenosine triphosphate (ATP) by ATP sulfurylase, and the amount of generated ATP fuels photon release, in proportional quantities, by luciferase. ATP and non-incorporated deoxynucleotides are removed by a washing step in the three-enzyme Pyrosequencing system of (1) (marketed by 454 life sciences) or degraded by apyrase in the four-enzyme system (2) (marketed by Biotage) (Fig. 1).

The four-enzyme system has been widely applied for single nucleotide polymorphism (SNP) genotyping (3-5), and to a lesser extent, for other applications such as the typing of bacteria (6), fungi (7), and viruses (8,9), determination of difficult secondary structures (10), mutation detection (11,12), DNA methylation analysis (13), multiplex sequencing (14), tag sequencing of a cDNA library (15,16), and clone checking (17).

On the other hand, the three-enzyme system has been used for whole bacterial genome sequencing (18), paleogenomics (19) and targeted deep sequencing of heterogeneous DNA material (20). These achievements were made possible by improving the Pyrosequencing chemistry and properly automating the processes. But the most significant improvements are the use of single-stranded DNA-binding protein (21), the purification of the most easily incorporated isomer of α -thio dATP in DNA synthesis (22) and the automation of a highly sensitive detection system (23). These changes allow routine the routine achievement of read-lengths of up to 100 nucleotides. Extended read-length will reduce the cost of DNA analysis for most applications. In addition, it will enable *de novo* sequencing of more complex eukaryotes. According to our recent research (24), mammalian genome sequencing can be assembled with high continuity by reading about 200 nucleotides, and total sequencing costs continue to decrease with longer reads. To pass the 100-nucleotide read-length barrier and reproducibly produce beyond 200-nucleotide reads, we sought to investigate potential limiting factors in Pyrosequencing's chemistry.

To find solutions for extending Pyrosequencing's read-length we experimentally characterized all the potential limitations in each single reaction involved in Pyrosequencing chemistry. Experimental kinetic data were then used to improve the Pyrosequencing enzyme model presented by Agah et al. (25). The extended model was then used to simulate the read-length of four-enzyme and three-enzyme Pyrosequencing under different conditions.

MATERIALS AND METHODS

Synthesis and purification of oligonucleotides

The oligonucleotide ROMO-Loop (5'-GCTGGAATTCGTCAGACTGGCCGTCGTTTTACAACGGAACGTTGTAAAACGACG G) was synthesized and purified by standard phosphoramidite chemistry with an in-house automated device at the Stanford Genome Technology Center.

Investigation of inhibitory products and dilution effects in Pyrosequencing

Pyrosequencing reactions were performed at room temperature in a volume of 50 μ l on the automated Pyrosequencing PSQTM96MA instrument (www.Pyrosequencing.com). The reaction mixture contained: 40 mU apyrase (Sigma Chemicals Co., St. Louis, MO, USA), 500 ng purified luciferase (www.Promega.com), 0.1 M Tris-acetate (pH 7.75), 0.5 mM EDTA, 5 mM magnesium acetate, 1 mM dithiothreitol, 0.4 mg/ml polyvinylpyrrolidone (360,000), and 100 μ g/ml D-luciferin (Sigma chemicals). All experiments were repeated five times and the

average data was calculated. To investigate product inhibition effects on ATP sulfurylase, luciferase, and apyrase in Pyrosequencing reactions, additional PPi and APS, AMP, ATP, or nucleotides were added to successive experiments, and the effect of these additions on sequence signals were analyzed. Specifically, to study the effect of product inhibition on a given enzyme (ATP sulfurylase for instance), the substrate of that enzyme was added to the reaction mixture, which was agitated for five minutes prior to the addition of nucleotides to the sequencing reaction. The light signal generated from nucleotide incorporation was then compared to light signal generated from standard Pyrosequencing reactions (with no additional substrate). A CCD camera was used to detect light output resulting from nucleotide incorporation in all experiments. The obtained data were analyzed using the SNP Evaluation tool (www.Pyrosequencing.com).

An additional experiment was performed to quantify the effects of dilution and evaporation. 50 μ l purified water was added to each of 10 different wells. Using the PSQTM96MA Pyrosequencing machine, 100 dispensations of 0.2 μ l of water were made into all wells over a period of 100 minutes. The volume of the water in all wells was measured immediately after the last dispensation.

Computer Simulations

Simulations were performed using MATLAB Version 7 (www.mathworks.com). Differential equations defined for each step in the Pyrosequencing reaction were simultaneously performed for all species as described previously (25). Experimentally obtained values and reactions for enzyme inhibitions were added to this model, and the Pyrosequencing reactions were simulated for up to 200 nucleotide dispensations on a 300 bp DNA fragment randomly chosen from human chromosome 1. The exact reactions added are shown in the Simulation Results section.

In order to study the factors inhibiting long Pyrosequencing read lengths, we further developed a model to simulate multiple dispensation cycles. This created some programming challenges, most notably, the generation of different lengths of DNA molecules resulting from unsynchronized extension. In addition, nucleotide incorporation for all DNA fragments present in the reaction had to be monitored. Therefore, arrays were allocated to store the concentration of different DNA fragments in various polymerization states, i.e. free, bound to DNA polymerase, in polymerase, DNA.dNTP complex, etc. Upon successful nucleotide incorporation, the DNA molecule is flagged as DNA(n+1). Thereby, the program determines potential incorporation at each stage by tracking PPi production from all possible DNA fragments.

In the three-enzyme system, apyrase is replaced with a washing step to eliminate excess nucleotides. Biotinylated DNA fragments are immobilized on streptavidin-coated paramagnetic beads (1). For each nucleotide dispensation cycle, the nucleotide and Pyrosequencing enzymes and reagents are added to each well containing the immobilized DNA. A washing step, wherein all other reagents but the DNA are removed, takes place every 30 seconds, between each nucleotide dispensation. Simulation was performed assuming 100%, 99.9%, 99%, and 90% washing efficiency. In addition, we presumed that 0.1% of DNA fragments are lost per wash.

RESULTS

Experimental Results

In order to determine factors limiting read-length in Pyrosequencing, we investigated the effect of product accumulation for each enzymatic reaction by varying reagent concentrations and studying sequence signal peak responses. Product accumulation decreases the catalytic

efficiency of the enzymatic reactions, thereby limiting Pyrosequencing read-length. The cascade of reactions is initiated by DNA polymerase, when a nucleotide complementary to the target strand is incorporated. No major factors inhibiting the DNA polymerase reaction have been observed when using natural nucleotides or the efficiently incorporated isomers of α -thio dATP, which is used in Pyrosequencing.

ATP sulfurylase

ATP sulfurylase inhibition by byproduct accumulation was investigated by adding different amounts of the substrates PPi and APS to the Pyrosequencing reaction mixtures. The sequence signals from these experiments were compared with standard Pyrosequencing. Five hundred, 1000, and 2000 pmol quantities of PPi and APS were added to three different reaction wells. As shown in Figure 2, signal intensities decrease by 20, 30, and 52 percent, respectively.

AMP

In similar experiments as described above, amounts of 500, 1000, and 2000 pmol of AMP were added to Pyrosequencing reaction mixtures, and the light signals observed were compared to a standard control. Figure 3 demonstrates the effects of adding these quantities of AMP to the reaction mixtures. As shown, addition of AMP over this concentration range has a minimal effect on sequence signals. The higher signal intensities shown in Figures 3b and 3c over the standard control were found to be due to instrument error, rather than the effect of AMP inhibition. This experiment was performed five times (data not shown), and the peak heights were not observed to have any inhibition correlation with AMP. Therefore, the results suggest that AMP is not a limiting factor in Pyrosequencing read length.

Oxyluciferin

To investigate the effects of possible oxyluciferin inhibition on luciferase, the same experiments as above were carried out but instead 500, 1000, and 2000 pmol of ATP were added to different reaction mixtures. The results of this experiment are presented in Figure 4. Compared to the control signal (Figure 4a), signal peaks were decreased by 5, 10, and 22%, respectively, when 500, 1000, and 2000 pmol of ATP were added.

Apyrase

To study the effect of byproduct accumulation on apyrase inhibition, signal quality was observed following the addition of varying nucleotide concentrations. The Pyrosequencer PSQ™96MA dispenses 0.2 μ l, or approximately 100 pmoles, of a given dNTP per cycle. In this experiment, five standard Pyrosequencing reactions were performed in parallel. The first sequence reaction contained DNA template as control, while the other four reaction mixtures did not contain any DNA template. Iterative nucleotide dispensations of 10, 20, 40, and 80 cycles were performed on wells two, three, four, and five, respectively. Subsequently, DNA template was added to each well. The effect of accumulated byproducts on apyrase activity was studied by observing the baseline broadness of signal peaks (the wider the baseline, the more inhibition) (Figure 5). Figure 5a illustrates signals from the standard Pyrosequencing reaction. Arrows highlight positions where nucleotide inhibition of apyrase can be observed. As shown, an increase in the number of cycles of nucleotide dispensations causes increasing byproduct inhibition of apyrase catalytic activity (Figure 5b-e).

Polymerase fidelity

Another potential factor limiting read-length is nucleotide misincorporation by DNA polymerase. To test the misincorporation rate of Klenow DNA polymerase, two reactions containing all standard Pyrosequencing reagents and enzymes except apyrase were examined. In one reaction well (Figure 6b), a mismatched nucleotide was dispensed. After 20 minutes of

observation, the correct nucleotide was dispensed to both solutions. The misincorporation rate for dGTP was calculated to be 0.17/1200 seconds (Figure 6) by comparing the height of the signal for correct incorporation and misincorporation within the same pyrogram.

Dilution effect

To measure the effects of dilution and evaporation, 100 dispensations of 0.2 μl of water were made in 10 different wells each of which initially contained 50 μl of water. Afterwards, the average volume in the wells was measured to be 54.0 $\mu\text{l} \pm 0.5 \mu\text{l}$. The volume of water in each well after 100 dispensations should have been 70 μl . This indicates that 22% (16 μl) of the reaction mixture evaporates over a period of 100 minutes. In other words, on average, the sequencing reaction mixture is being diluted 0.07% (or 0.04 μl) after each nucleotide dispensation, and this dilution phenomenon will gradually affect the concentration balance of enzymes and reagents in the Pyrosequencing reaction.

Simulation Results

Based on the obtained results from the above-mentioned experiments, the following reactions were added to the previously described Pyrosequencing model (25) to represent the inhibition effects of byproduct accumulation in Pyrosequencing reactions.

Furthermore, the effect of dilution on each enzyme after each cycle was incorporated into the proposed simulation model.

Figure 7 demonstrates the result of simulating the final four-enzyme model for 150 nucleotide dispensations. After approximately 80 dispensations, the signal-to-noise ratio decreases until it becomes more difficult to distinguish signals and noise as well as single, double, and triple base signal peaks. If only the sequencing data from the first 80 nucleotide dispensations are considered, an accurate base-calling of about 60 bases can be obtained. This result is in alignment with previously reported experimental results (22). However, more promising results were obtained from simulating the three-enzyme Pyrosequencing reaction (see below).

Next, we examined the effects of manual washing for nucleotide removal, as opposed to enzymatic degradation. To study the washing efficiency in the three-enzyme system, 100% and 90% washing efficiencies were performed in 200 nucleotide dispensations. Figure 8a and 8b demonstrate that even 90% washing efficiency is able to generate sequencing reads of more than 400 nucleotides. Figure 9 presents the simulation results of the 90% washing efficiency system for 300 nucleotide dispensation. Noise in the lower panel of Figure 9 remains insignificant even after 250 nucleotide dispensations. Three-enzyme Pyrosequencing with 99% and 99.9% washing efficiencies produced very similar data to 100% washing efficiency (data not shown). It is worth noting that the decrease in signal intensity (evident in the top portion of Figure 9) results from the assumption that 0.1% of DNA fragments are lost during each washing cycle. These simulation results point to the ability of the three-enzyme system to generate much longer read lengths.

DISCUSSION

Pyrosequencing employs a cascade of enzymes to provide sequence data for DNA fragments obtained via PCR. The activity of the enzymes involved can be observed in Pyrosequencing signal peaks. The slope of the ascending curve relative to the peak point demonstrates the activity of the DNA polymerase and ATP sulfurylase. The height of the signal is determined by the activity of luciferase. The slope of the descending curve demonstrates the apyrase

activity in the four-enzyme system, and washing efficiency in the three-enzyme system. In order to systematically address the issues in long read Pyrosequencing, we investigated the kinetic properties of each enzyme separately, and characterized potential limiting factors. A new simulation model was set up to study the effects of changes in various parameters in obtaining long reads. Experimental and simulation results of three-enzyme and four-enzyme systems are discussed here separately.

Four-enzyme system of Pyrosequencing

Routinely, about 60 bases can be obtained on most DNA templates using the commercial instruments PSQTM96MA and HSTM96. However, sequence reads of more than 200 bases have been reported (22). Each enzyme involved in this system was investigated separately. The DNA polymerase reaction results in two products, extended DNA and PPi. DNA polymerase has shown higher affinity to double-stranded DNA than single-stranded DNA (26). Although nascent double-stranded DNA demonstrates higher DNA polymerase affinity, we have not seen any noticeable inhibition of DNA polymerase. This may be due to the fact that the Pyrosequencing system uses 5 to 10 times more DNA polymerase than DNA fragments. The PPi released during polymerization is fully converted to ATP by ATP sulfurylase; therefore, no inhibition is expected. On the other hand, unincorporated nucleotides are degraded to dNMP's by apyrase. These byproducts may inhibit DNA polymerase, based on comparisons of simulation and experimental results. A bidirectional reaction in which dNMPs complex with DNA polymerase was added to our simulation model to coordinate simulation and experimental results. Accumulation of dNMPs was also found to affect apyrase efficiency as well.

Another potential factor limiting read-length is nucleotide misincorporation by DNA polymerase. In the three-enzyme Pyrosequencing system, the rate of misincorporation was calculated to be 0.015% per second. Naturally, this rate would be far lower in the four-enzyme system due to immediate degradation of excess nucleotides by apyrase, which drives nucleotide concentration below the K_M value of DNA polymerase in a few seconds. However, since the rate of subsequent correct incorporation following misincorporation is very low (27), we can reasonably assume that those DNA fragments with misincorporated nucleotides are nonexistent in later cycles. Therefore, if only misincorporation occurred, the light signal would decrease at a constant rate, but unlike non-synchronized DNA extension, the noise would stay constant. With the rate of 0.9% misincorporations per cycle of one minute, the intensity of light signals due to correct incorporation drops by 50% after 80 unmatched nucleotide dispensations in the three-enzyme system. To limit the effect of misincorporation on Pyrosequencing's read-length a polymerase with lower misincorporation rate is recommended. Furthermore, shorter cycle durations can be utilized to minimize the presence of unmatched nucleotides for incorporation into DNA fragments.

Another potential factor limiting the read-length may be effect of SO_4^{2-} on ATP sulfurylase. Various amounts of PPi and APS were added to different wells and signal peaks were compared to the control. Note that the luciferase and apyrase in the solutions consume all available ATP. Therefore, it was assumed that the ATP sulfurylase reaction would not reach equilibrium, and all the added PPi and APS would be converted to ATP and SO_4^{2-} . As shown in Figure 2, Pyrosequencing signal heights drop by as much as 52%, when 2000 pmol of PPi and APS are added. However, this decrease is partially due to ATP sulfurylase inhibition by SO_4^{2-} . Accumulation of SO_4^{2-} slows the forward reaction of ATP sulfurylase and hence the rate of ATP production is hampered, thereby decreasing signal intensity. We conclude that accumulation of SO_4^{2-} is directly correlated with a decrease in the peak of Pyrosequencing signals. Furthermore, addition of PPi and APS to the Pyrosequencing solution leads to the accumulation of products such as AMP and oxyluciferin as well as SO_4^{2-} . The inhibitory effects

of AMP and oxyluciferin were studied separately to distinguish inhibitory effects of SO_4^{2-} alone on Pyrosequencing signals.

The inhibition effects of product accumulation on luciferase were also studied. Various amounts of AMP and ATP were added to different solutions and signal peaks were compared to the control. Note that ATP addition was performed to study the effects of oxyluciferin on luciferase activity. It was found that the addition of even 2000 pmol of AMP did not significantly affect signal peaks (Figure 3). However, addition of ATP, and hence accumulation of oxyluciferin, resulted in decreased Pyrosequencing signal peak values by as much as 22% (when 2000 pmol ATP was added) (Figure 4). These results, combined with those obtained from ATP sulfurylase inhibition experiments, suggest that the addition of 2000 pmol SO_4^{2-} reduces signal heights by approximately 30%. Introduction of 2000 pmol of PPi, ATP, or AMP is essentially equivalent to 2000 base incorporations, since 1 pmol of DNA is used in the four-enzyme system. Thus, if ATP sulfurylase and luciferase product inhibition were the only read length limiting factors, the read-length should have been much greater than current 60 bases. This suggests that ATP sulfurylase and luciferase inhibition are not the main limiting factors.

To study the effect of nucleotide byproduct accumulations on apyrase catalytic activity, different numbers of nucleotide dispensations were performed. As highlighted with arrows in Figure 5, dNMP and dNDP (byproducts of nucleotide degradation by apyrase) accumulation has two visible effects on signal peaks. These effects are due to apyrase inefficiency in degrading excess nucleotides and ATP following each dNTP addition. This inefficiency is represented in the broadening and height decrease of signal peaks, and the duration of descending curves. As more nucleotides are added to solutions prior to standard Pyrosequencing, signals are diminished by apyrase at a much slower rate (Figure 5). Beyond 20 nucleotide dispensations, signal intensity does not reach zero within 60 seconds, the standard cycle duration. Decreased apyrase activity results in nucleotide accumulation and causes asynchronous extension as well as non-uniform peaks in later cycles of Pyrosequencing. Asynchronous DNA extension is a potential limitation in Pyrosequencing as it decreases the intensities of correct signals and increases background signals. The decrease in signal intensities is clear when comparing the sequence signals in Figure 5b-e to the control signals (Figure 5a). Vertical arrows in Figure 5 highlight the occurrence of noise signals in wells containing additional nucleotides prior to Pyrosequencing reactions. Based on these results, it is believed that apyrase inefficiency in degrading excess nucleotides in later cycles is the main factor constricting the read length of a four-enzyme Pyrosequencing system.

Based on the experimental results above, the following two reactions were added to the four-enzyme Pyrosequencing system model presented by Agah and colleagues (25):

These reactions take the inhibitory effects of dNMP and dNDP on apyrase and polymerase into account. In our simulations, we considered dNMP and dNDP as the same byproduct for simplicity. Moreover, the inhibitory effects of other byproducts such as oxyluciferin were accounted for in the model presented previously (25), as reactions involving those products are considered to be reversible and bidirectional. The simulation result of the four-enzyme system is presented in Figure 7. Interestingly, the error-free base-calling achieved for this model is approximately 60 bases, which agrees with read-lengths obtained using the commercial Pyrosequencing machine PSQTM96MA. Furthermore, the apyrase inhibition becomes more apparent after the 50th nucleotide dispensation, accompanied by an increase in the intensity of noise signals. Beyond 80th nucleotide dispensation, the quality of base-calling decreases significantly.

Simulation and experimental results both highlight the importance of apyrase efficiency in degrading excess nucleotides for a longer read-length. To reiterate, the excess nucleotides accumulated from Pyrosequencing cycles cause non-synchronized DNA extension, which decreases signal intensities and increases noise. The signal-to-noise ratio decreases until error-free sequencing becomes impossible. To increase the read-length of four-enzyme system, we suggest two possible solutions. The first is to enhance enzymatic nucleotide removal efficiency in degrading excess nucleotides. The apyrase used in the four-enzyme system Pyrosequencing is obtained from *Solanum tuberosum*, which demonstrates 90% higher efficiency in degrading dNTP to dNDP than dNDP to dNMP (28). Thus, the main product inhibition of apyrase is due to accumulation of dNDP in the solution, rather than dNMP. Adding a small amount of dNDP- and dNMP-degrading enzymes would potentially increase the efficiency of nucleotide degradation, thereby allowing longer reads. A second solution for increasing the read-length of the four-enzyme system is replacing apyrase with a washing step. Any product inhibition can be avoided by using washing to remove accumulated byproducts and excess nucleotides. This solution is used in the three-enzyme system of Pyrosequencing.

Three-enzyme system of Pyrosequencing

In the three-enzyme system, enzymatic nucleotide removal by apyrase is replaced with washing steps. The other enzymatic reactions remain unchanged. In such a system, DNA fragments are immobilized and capable of being captured and removed from the system. New enzymes and reagents are added after each washing step. Average read-lengths of 106 nucleotides are routinely achieved by Genome Sequencer 20 (www.454.com) on various DNA samples (29). To further investigate the three-enzyme system, we inserted all the obtained inhibitory values from individual enzymatic reactions and set up a new simulation model to study the read-length of Pyrosequencing given 100%, 99.9%, 99%, and 90% washing efficiencies. Each nucleotide dispensation cycle in the three-enzyme system (available through 454 Life Sciences) takes about 90 seconds comprising of ~ 20 seconds of nucleotide flow and ~ 70 seconds of washing. In our model, we assumed 5 seconds for a hypothetical washing step. According to our simulation results (Figure 8 and 9), with even 90% washing efficiency read-lengths over 300 bases could be achieved. The nucleotide removal efficiency through washing can be enhanced by inclusion of apyrase enzyme in the wash buffer, which is already performed in Genome Sequencer 20 system from 454 Life Sciences. We believe that there is still room for improvement in this step which is critical for increasing the read length of Pyrosequencing.

The washing step in the three-enzyme system results in removal of all the inhibiting byproducts along with enzymes and reagents. The washing step causes a loss of DNA fragments and necessitates addition of new enzymes and reagents, which increase the overall expense of the process. The use of advanced microfluidic systems that miniaturize the Pyrosequencing reaction could significantly reduce the total costs of the three-enzyme system. While 454's Genome Sequencer has partly addressed this, each nucleotide feed is still about 2 milliliters. According to our simulation data, this volume could potentially be reduced by 1000-fold (data not shown). The problem of DNA loss as a result of extensive washing could potentially be addressed by more stable immobilization schemes and trapping the beads within the wells of a picotiter plate by different means. However, this is not a major problem for mammalian genome sequencing, as even a 0.1% of DNA loss after each wash, as simulated with our model, still produces over 300 nucleotides of sequence data (Figure 9) and further improvement is envisioned in the chemistry and mechanical aspect of the system which may increase the read-length even beyond 500 bases. This read-length is likely to be necessary for high quality genome assembly using direct shotgun sequencing of mammalian genomes.

CONCLUSION

The major factor limiting the read-length obtained via Pyrosequencing is non-synchronized extension of DNA fragments. Although the read-length in the four-enzyme system could be improved by enhanced nucleotide-removal efficiency, byproduct accumulation will still limit the system. Here, we have demonstrated that longer reads can be achieved via three-enzyme Pyrosequencing, wherein apyrase is excluded from the sequencing system and byproducts are removed with a washing step (Table 1). Detailed simulation analysis indicates the potential for read-lengths well beyond 300 bases. To obtain longer reads, improved software for base-calling may be implemented. In addition, the use of homogenous fragment lengths in the DNA shearing step of clonal amplification would provide a more economical scheme for DNA sequencing - it is too costly to continue sequencing if the majority of DNA fragments are fully extended in a sequencing run. It is worth noting that enhanced detection sensitivity would provide a higher signal-to-noise ratio, thereby allowing a highly miniaturized system and more cost-effective DNA sequencing.

Acknowledgements

The authors are supported by NIH grant # R01HG003571, and PO1HG00205. We would like to thank Dr. Baback Gharizadeh, Ali Agah, Ronald W. Davis, Peter Griffin and Dr. Mohsen Nemat-Gorgani for useful discussions.

References

- Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry* 1996;242:84–89. [PubMed: 8923969]
- Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. *Science* 1998;281:363. [PubMed: 9705713]
- Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P, Uhlen M, Lundeberg J. Single-nucleotide polymorphism analysis by Pyrosequencing. *Anal Biochem* 2000;280:103–110. [PubMed: 10805527]
- Fakhrai-Rad H, Pourmand N, Ronaghi M. Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. *Hum Mutat* 2002;19:479–485. [PubMed: 11968080]
- Ronaghi M. Pyrosequencing for SNP genotyping. *Methods Mol Biol* 2003;212:189–195. [PubMed: 12491911]
- Ronaghi M, Elahi E. Pyrosequencing for microbial typing. *J Chromatogr B Analyt Technol Biomed Life Sci* 2002;782:67–72.
- Gharizadeh B, Norberg E, Loffler J, Jalal S, Tollemar J, Einsele H, Klingspor L, Nyren P. Identification of medically important fungi by the Pyrosequencing technology. *Mycoses* 2004;47:29–33. [PubMed: 14998396]
- Elahi E, Pourmand N, Chaung R, Rofoogaran A, Boisver J, Samimi-Rad K, Davis RW, Ronaghi M. Determination of hepatitis C virus genotype by Pyrosequencing. *J Virol Methods* 2003;109:171–176. [PubMed: 12711060]
- Gharizadeh B, Kalantari M, Garcia CA, Johansson B, Nyren P. Typing of human papillomavirus by pyrosequencing. *Lab Invest* 2001;81:673–679. [PubMed: 11351039]
- Ronaghi M, Nygren M, Lundeberg J, Nyren P. Analyses of secondary structures in DNA by pyrosequencing. *Analytical Biochemistry* 1999;267:65–71. [PubMed: 9918656]
- Ahmadian A, Lundeberg J, Nyren P, Uhlen M, Ronaghi M. Analysis of the p53 tumor suppressor gene by pyrosequencing. *Biotechniques* 2000;28:140–144. [PubMed: 10649785]
- Garcia AC, Ahmadian A, Gharizadeh B, Lundeberg J, Ronaghi M, Nyren P. Mutation detection by Pyrosequencing: sequencing of exons 5 to 8 of the p53 tumour suppressor gene. *Gene* 2000;253:249–257. [PubMed: 10940563]
- Uhlmann K, Brinckmann A, Toliat MR, Ritter H, Nurnberg P. Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* 2002;23:4072–4079. [PubMed: 12481262]

14. Pourmand N, Elahi E, Davis RW, Ronaghi M. Multiplex Pyrosequencing. *Nucleic Acids Res* 2002;30:e31. [PubMed: 11917037]
15. Gharizadeh B, Herman ZS, Eason RG, Jejelowo O, Pourmand N. Large-scale Pyrosequencing of synthetic DNA: A comparison with results from Sanger dideoxy sequencing. *Electrophoresis* 2006;27:3042–3047. [PubMed: 16800029]
16. Nordstrom T, Gharizadeh B, Pourmand N, Nyren P, Ronaghi M. Method enabling fast partial sequencing of cDNA clones. *Anal Biochem* 2001;292:266–271. [PubMed: 11355860]
17. Nourizad N, Gharizadeh B, Nyren P. Method for clone checking. *Electrophoresis* 2003;24:1712–1715. [PubMed: 12783446]
18. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437:376–380. [PubMed: 16056220]
19. Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, et al. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* 2006;311:392–394. [PubMed: 16368896]
20. Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander EC Jr, Rohwer F. Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* 2006;7:57. [PubMed: 16549033]
21. Ronaghi M. Improved performance of Pyrosequencing using single-stranded DNA-binding protein. *Anal Biochem* 2000;286:282–288. [PubMed: 11067751]
22. Gharizadeh B, Nordstrom T, Ahmadian A, Ronaghi M, Nyren P. Long-read pyrosequencing using pure 2'-deoxyadenosine-5'-O'-(1-thiotriphosphate) Sp-isomer. *Anal Biochem* 2002;301:82–90. [PubMed: 11811970]
23. Ronaghi M. Pyrosequencing sheds light on DNA sequencing. *Genome Res* 2001;11:3–11. [PubMed: 11156611]
24. Sundquist A, Ronaghi M, Tang H, Pevzner P, Batzoglou S. Whole-Genome Sequencing and Assembly with High-Throughput, Short-Read Technologies. 2006In Press
25. Agah A, Aghajan M, Mashayekhi F, Amini S, Davis RW, Plummer JD, Ronaghi M, Griffin PB. A multi-enzyme model for Pyrosequencing. *Nucleic Acids Res* 2004;32:e166. [PubMed: 15576673]
26. Ljach MV, Kolocheva TI, Gorn VV, Levina AS, Nevinsky GA. The affinity of the Klenow fragment of *E. coli* DNA-polymerase 1 to primers containing bases noncomplementary to the template and hairpin-like elements. *FEBS Lett* 1992;300:18–20. [PubMed: 1547884]
27. Nyren P, Karamohamed S, Ronaghi M. Detection of single-base changes using a bioluminometric primer extension assay. *Analytical Biochemistry* 1997;244:367–373. [PubMed: 9025954]
28. Traverso-Cori A, Chaimovich H, Cori O. Kinetic studies and properties of potato apyrase. *Arch Biochem Biophys* 1965;109:173–181. [PubMed: 14281943]
29. Margulies M, Egholm M, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437:376–380. [PubMed: 16056220]

Abbreviations

PPi	Inorganic Pyrophosphate
Pi	Inorganic Phosphate
ATP	Adenosine triphosphate
AMP	Adenosine monophosphate
ADP	Adenosine diphosphate

APS	Adenosine phosphosulfate
dNTP	2'-deoxy nucleotide-5'-tri phosphate
dNMP	2'-deoxy nucleotide-5'-monophosphate
dNDP	2'-deoxy nucleotide-5'-diphosphate

Figure 1.
Enzymatic reactions of the four-enzyme Pyrosequencing system.

Figure 2. Effect of product inhibition on ATP sulfurylase

(a) signals from the control solution. Amounts of 500 pmol (b), 1000 pmol (c), and 2000 pmol (d) of PPi and APS were dispensed into three different solutions. The signal intensities decrease by 20%, 30%, and 52% in (b), (c), and (d) respectively as compared to signal intensities in (a).

Figure 3. Effect of AMP inhibition on luciferase

(a) control solution 500 pmol, 1000 pmol, and 2000 pmol of AMP were added to solution in (b), (c), and (d) respectively. The light signals' peak heights and shapes do not change significantly in these solutions compared to the signal from the control solution.

Figure 4. Effect of oxyluciferin inhibition on luciferase

(a) is the control solution. 500 pmol, 1000 pmol, and 2000 pmol of ATP were added to (b), (c), and (d) respectively prior to Pyrosequencing. Signal peaks decreased by 5, 10, and 22% in (b), (c), and (d) respectively compared to the control signal.

Figure 5. Effect of product inhibition on pyrophosphatase

(a) is the control solution. DNA fragments were added to (b) after 10, to (c) after 20, to (d) after 40, and to (e) after 80 dNTP dispensations. More noise (vertical arrows) due to unsynchronized DNA extension is visible in solutions with more dNTP dispensations.

Figure 6. Effect of misincorporation by DNA polymerase on Pyrosequencing signals

(a) is the control solution; the solution in (b) was incubated with an unmatched nucleotide for 20 minutes (in the absence of apyrase) before the complementary nucleotide was dispensed. The light signal in (b) is 17% less than the control signal.

Figure 7. Simulation results of four-enzyme Pyrosequencing system on a 300-base long DNA fragment. Error-free base-calling is achieved for 60 bases in this simulation result.

Figure 8. Simulation results of three-enzyme system Pyrosequencing on a 300-base long DNA fragment with (a) 100% and (b) 90% washing efficiency. Noise remains minimal in both cases resulting in much longer error-free read-length compared to four-enzyme Pyrosequencing system.

Figure 9. Simulation results of the three-enzyme Pyrosequencing system on the same DNA fragment as above, but with 300 nucleotide dispensations. The signal intensity decreases slightly over 300 nucleotide dispensations (top); even during the later cycles and by 300th dispensation, signal-to-noise ratio remains acceptable for an error free base-calling (bottom). This shows major improvement compared to simulation result obtained for the four-enzyme system.

Table 1

Comparison of four-enzyme and three-enzyme Pyrosequencing systems

	Pros	Cons	Primary limiting factor	Increasing read-length
Four-enzyme system	<ul style="list-style-type: none"> - More easily automated - No need to have a washing step 	Limited read length	<ul style="list-style-type: none"> - Non-synchronized DNA extension due to enzyme denaturation, inhibition, or dilution (due to consecutive nucleotide dispensation), specially of apyrase - Misincorporation and incomplete DNA extension 	<ul style="list-style-type: none"> - Use more stable and efficient enzymes - Add new enzymes after 50 cycles
Three- enzyme system	Longer read- length	Harder to automate sequencing due to need for a washing step	<ul style="list-style-type: none"> - Lost of DNA templates - Incomplete DNA extension 	<ul style="list-style-type: none"> - Minimize the duration of the washing step - Use novel method for DNA immobilization