

# DNA sequencing using the liquid polymer POP-7™ on an ABI PRISM® 3100 Genetic Analyzer

Michelle M. Detwiler, Timothy J. Hamp, and A. Latif Kazim  
Roswell Park Cancer Institute, Buffalo, NY

*BioTechniques* 36:932-933 (June 2004)

A common platform for fluorescent dideoxy DNA sequencing (1) is the capillary-based automated DNA sequencer (2–4). Capillary-based sequencing uses liquid self-coating polymers (5) to separate DNA fragments and has vastly improved the efficiency of DNA sequencing. Many core facilities use the Applied Biosystems ABI PRISM® 3100 Genetic Analyzer, a 16-capillary automated DNA sequencer. Several capillary lengths and protocols are supported by the manufacturer for DNA sequencing (ABI PRISM 3100 Genetic Analyzer, User's Manual). These protocols include the use of POP-6™ polymer with 50-cm arrays, as well as POP-4™ polymer with 80-cm arrays. The 50-cm array/POP-6 protocol has shorter run times but may only yield 650–700 high-quality bases. The 80-cm array/POP-4 protocol can generate read lengths of 900 bases, but with longer run times (approximately 3.5 h). POP-7™, the newest polymer developed specifically for the Applied Biosystems model 3100 Genetic Analyzer, has been shown to combine long read lengths with shorter run times but is not supported for use on the 3100 Genetic Analyzer. We have developed a modified run protocol that utilizes POP-7 and 50-cm arrays on the 3100 Genetic Analyzer. With this protocol, we are able to obtain accurate sequencing of 850–900 bases in 2 h, 5 min. Thus, our method provides increased sequencing throughput in comparison with the other supported long-read sequencing protocols without compromising high-quality basecalling and read length.

Data was obtained using the BigDye® Terminator 3.1 Sequencing Standard. POP-4, POP-6, and POP-7 Per-

formance Optimized Polymers were used as the separation matrices. A 10× Genetic Analysis buffer with EDTA, diluted to 1×, was used as the running buffer for POP-4 and POP-6 runs. 3730 Buffer (10×) with EDTA, diluted to 1×, was used for POP-7 runs. All reagents, standards, software programs, and updates were from Applied Biosystems (Foster City, CA, USA).

Sequence data was collected using ABI PRISM 3100 Data Collection™ v1.1. Our POP-7 sequencing protocol uses run parameter values similar to those in the standard POP-6/50-cm array-sequencing module (provided in Data Collection v1.1), but with specific modifications for improved performance. Our alterations are listed in Table 1.

Data were analyzed with ABI PRISM DNA Sequencing Analysis Software™ v5.1, upgraded with the software patch KB™ v1.1.1. This upgrade was necessary to address an issue in which early truncation of the basecalling analysis occurred, resulting in shorter than expected read lengths. This effect was seen when the original KB basecaller v1.1, included in the Sequencing Analysis v5.1 software, was applied to samples that migrated through the capillaries differently than was expected by the basecaller calibration data. This effect generally occurs when standard run parameters, such as temperature or voltage, are modified (Release Notes on KB basecaller v1.1.1 software update). Sequencing Analysis v5.1 contains all mobility files and basecallers necessary to analyze the data, a feature that was key to the success of our protocol as it is necessary to use the KB basecaller in conjunction with POP-7 mobility files to achieve the highest quality data.

To assess the performance and reliability of this modified protocol, we used the BigDye v3.1 Sequencing Standard for our experimental runs. We aligned our acquired data against a BigDye reference sequence, using the Seq Autoscore v2.0 program to evaluate the overall accuracy. We compared our

**Table 1. Comparison of Applicable Run Parameters**

Run Parameter	Standard ABI POP-6 Module	Module Modified for POP-7
Run Temperature (°C)	50.0	60.0
Pre-Run Voltage (kV)	12.2	8.5
Run Voltage (kV)	12.2	8.5
Data Delay Time (s)	1200.0	240.0
Run Time (s)	6500.0	6200.0
Injection Time (s)	22.0	40.0

**Table 2. Accuracy Values and KB QV20 Scores of BigDye v3.1 Sequencing Standards as Run with Various Polymer/Array Combinations**

Accuracy (bases)	POP-4 80-cm Array (%)	POP-6 50-cm Array (%)	POP-7 50-cm Array (%)
800	99.94	99.97	99.99
900	99.94	99.30	99.99
1000	99.86	97.62	99.89
<b>Average KB basecaller QV20 Score (bp)</b>	755	694	798

All samples were analyzed with the KB basecaller and appropriate mobility files using Sequencing Analysis v5.1. Accuracy was determined by comparing samples to a BigDye standard reference sequence using the Seq Autoscore v2.0 program. Quality value (QV) scoring was applied by the KB basecaller.

average read length accuracy values to those obtained using the other two long-read sequencing protocols supported by the manufacturer. In addition, per-base quality value scores (QV20) were generated by the KB basecaller. Our findings are summarized in Table 2.

The advantages of incorporating this protocol are numerous. First, throughput can be substantially increased in comparison to other supported long-read sequencing protocols. When using a POP-4/80-cm array protocol, 96 samples can be processed in 24 h, while the POP-6/50-cm array method has a 24-h throughput of 144 samples. In contrast, our modified POP-7 protocol, with its faster run times, provides sequence data for 176 samples per day. Second, the polymer is one of the most expensive consumables associated with capillary-based sequencing. POP-7 is approximately one-fourth the cost of either POP-4 or POP-6, and its use can significantly improve the cost-effectiveness of sequencing with the 3100 Genetic Analyzer. Third, in our experience, POP-7 appears to have a lower incidence of artifactual "spikes" in sequencing data. Spikes appear as wide, multicolored peaks that can obscure nucleotides and are thought to be caused by microbubbles or particles within the polymer that pass by the detector. The KB basecaller included in Sequencing Analysis v5.1 has a built-in "despiking" algorithm (<http://docs.applied-biosystems.com/pebiiodocs/00106525.pdf>); this algorithm can often subtract these spikes, but some are too large to be accurately corrected. With POP-7, the overall occurrence of these artifacts appears to be greatly reduced, with a subsequent decline in the number of reruns that may have to be performed to clarify affected samples.

In summary, our modified POP-7 sequencing protocol provides a practical alternative for DNA sequencing facilities wishing to maximize the performance of their 3100 Genetic Analyzer. Using the latest polymer, POP-7, in conjunction with the most recent version of Sequencing Analysis software (v5.1), we have been able to increase throughput, reduce operating costs, and improve the overall efficiency of our 3100 Genetic Analyzers.

## ACKNOWLEDGMENTS

*This work was supported, in part, by the Cancer Center Support grant no. CA016056-27.*

## REFERENCES

1. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
2. Luckey, J.A., H. Drossman, A.J. Kostichka, D.A. Mead, J. D'Cunha, T.B. Norris, and L.M. Smith. 1990. High speed DNA sequencing by capillary electrophoresis. *Nucleic Acids Res.* 18:4417-4421.
3. Cohen A.S., D.R. Najarian, and B.L. Karger. 1990. Separation and analysis of DNA sequence reaction products by capillary gel electrophoresis. *J. Chromatogr.* 516:49-60.
4. Swerdlow H., S.L. Wu, H. Harke, and N.J. Dovichi. 1990. Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. *J. Chromatogr.* 516:61-67.
5. Barbier, V. and J.L. Viovy. 2003. Advanced polymers for DNA separation. *Curr. Opin. Biotechnol.* 14:51-57.

Received 24 March 2004; accepted 12 April 2004.

*Address correspondence to A. Latif Kazim, Biopolymer Core Resource, Biophysics and Biochemistry Program, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA. e-mail: Latif.kazim@roswellpark.org*

## Procedure for controlling number of repeats, orientation, and order during cloning of oligonucleotides

Eran Blachinsky, Irit Marbach, Ruth Cohen, Melanie R. Grably, and David Engelberg

*The Hebrew University of Jerusalem, Jerusalem, Israel*

*BioTechniques* 36:933-936 (June 2004)

Short DNA sequences play powerful roles in a variety of experimental and industrial systems. Some sequences serve, for example, as binding sites for transcriptional activators or suppressors (1-3); others may encode short peptides that serve as tags on recombinant proteins (4); and yet others, encoding recognition sites for restriction enzymes, create multicloning sites.

Introduction of such sequences into plasmids is commonly used to create reporter genes for transcriptional research, for expression and purification of recombinant proteins, and for other purposes. In many cases, a particular number of repeats in a particular orientation is preferable or even required. However, no straightforward, reliable method for controlling the number and orientation of those short sequences is currently available.

Some attempts were made to develop systematic methods for controlling the insertions of DNA in tandem repeats (5-8). A basic useful method is to design the desired oligonucleotide carrying the short sequence (the insert) with flanking-compatible cohesive ends (7). These oligonucleotides are then self-ligated to give rise to a series of DNA fragments, each with a random copy number and orientation of the insert. Different tricks could be applied to confirm the orientation of the inserts within the DNA fragments or to enforce a directional cloning (7). Then, the self-ligation products are separated by gel electrophoresis and the chosen insert length is excised, purified, and ligated into the vector. This approach is quite useful for obtaining tandem repeats of the same oligonucleotide in the desired orientation, but it does