Benchmarks

s at 95°C followed by 30 s at 70°C. The sequencing gel was made of 8% Long Ranger™ gel solution (BMA, Rockland, ME, USA) and containing 7 M urea (Research Organics, Cleveland, OH, USA) and 1× TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA·Na₂ [pH 8.4]). Three microliters of formamide loading dye (from the Amersham Pharmacia Biotech kit) were added to the samples after the cycle sequencing reaction. The loading membranes (sample loader) were from Genetic BioSystems (San Diego, CA, USA). The samples were heated for 2.5 min at 95°C and either immediately spotted onto the tabs of a preheated (95°C) membrane loader (hot sample application) or cooled down in ice-cold water bath before spotting (regular sample application).

REFERENCES

1. Cassel, S.M. and A. Guttman. 1998. Membrane-mediated sample loading for automated DNA sequencing. *Electrophoresis* 19:1341-1346.
2. Guttman, A., C. Barta, M. Szoke, M. Sasvari-Szekely and H. Kalasz. 1998. Real-time detection of allele-specific polymerase chain reaction products by automated ultra-thin-layer agarose gel electrophoresis. *J. Chromatogr. A* 828:481-487.
3. Griffin, H.G. and A.M. Griffin. (Eds.) 1993. *DNA Sequencing Protocols: Methods in Molecular Biology*, Vol 23, Humana Press, Totowa, NJ.
4. Mizusawa S., S. Nishimura and F. Seela. 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucleic Acids Res.* 14:1319-1324.
5. Ruiz-Martinez, M.C., J. Berka, A. Belenkii, F. Foret, A.W. Miller and B.L. Karger. 1993. DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection. *Anal. Chem.* 65:2851-2858.
6. Tabor, S. and C.C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* 84:4767-4771.
7. Yamakawa, H., D. Nakajima and O. Ohara. 1996. Identification of sequence motifs causing band compressions on human cDNA sequencing. *DNA Res.* 3:81-86.

This work was supported by the US-Hungarian joint fund (J.F. No. 96/654) and by the Hungarian grants FKFP 0658/1999 and OTKA T022608. Address correspondence to Dr. Andras Guttman, Novartis

Agricultural Discovery Institute, 3115 Merryfield Row, La Jolla, CA 92121, USA. Internet: andras.guttman@nadii.novartis.com

Received 12 October 1999; accepted 27 December 1999.

Arpad Gerstner, Maria Sasvari-Szekely, Huba Kalasz and Andras Guttman¹

Semmelweis University

Medical School

Budapest, Hungary

¹Novartis Agricultural

Discovery Institute

La Jolla, CA, USA

Mix of Sequencing Technologies for Sequence Closure: An Example

BioTechniques 28:630-634 (April 2000)

There are many methods to overcome troublesome regions for sequence finishing, including primer walking, human editing, re-cloning and variations in cycle sequencing parameters, polymerases and other approaches such as reverting to radiolabeled sequencing or the LI-COR DNA sequencer. These methods are used daily by technicians to close sequence gaps between contigs as part of the Human Genome Project, each with varying levels of success, and none is guaranteed to work universally (1,3,5,7). With the Human Genome Project focused on the rapid accumulation of "draft sequence" and the ever challenging cloning and sequencing of telemeric regions, unfinished sequences are accumulating and highlight the need for alternative approaches. Indeed, researchers with a particular interest in a sequence region for which only draft sequence is available may want to finish it themselves and therefore must try all available finishing approaches.

Here, we describe the completion of a region of human genome sequence that was problematic and could only be closed through the use of an alternative gel-based technology using its specific and unique chemistry. A region of 602 bp, incorporating an Alu repeat and flanked by simple repeat sequences, could not be closed using all available ABI 377 chemistries alone (PE Biosystems, Foster City, CA, USA), even after many attempts (4). This was due to "hard stops" (frequently attributable to secondary structures or contaminants in the DNA sample) in the sequence, with an abrupt loss of signal as a result of incomplete polymerase extension through these structures. Using another sequencing technology, the new CEQ 2000 capillary sequencer (Beckman Coulter, Fullerton, CA, USA), with its novel infrared dye set and associated protocol, immediately produced results (technical information document available at www.beckman.com/Beckman/biorch/prodinfo/autodna/pdf/t1854a).

Benchmarks

pdf). These experiments suggest that a combination of different sequencing chemistry approaches enables effective completion of difficult regions that may be susceptible to polymerase stalling.

There were many attempts to close the 67 bp gap using (i) two different walking primers from each side and several specific primers designed from the partial known sequence; (ii) amplified fragments and subcloned fragments as targets; (iii) available chemistries (BigDye™, dRhodamine and dGTP substitutions for dITP in the BigDye kits; all from PE Biosystems) and (iv) a variety of adjustments to the cycle sequencing parameters on an ABI 377 sequencer, including annealing temperature adjustments and the addition of a betaine compound (6).

The typical sequencing method on the ABI 377 sequencer was as follows: 200 ng/μL DNA, 3.2 pmol primer, cycle sequencing reagent (AmpliTaQ® DNA polymerase, FS; PE Biosystems) and distilled water were combined and amplified at 96°C for 30 s, 50°C for 30 s and 60°C for 3 min, for 35 cycles. The excess dye was removed with a 96-well block spin column (Edge BioSystems, Gaithersburg, MD, USA). The samples were dried in a SpeedVac® (Savant Instruments, Holbrook, NY, USA) for 50 min and then resuspended in formamide. The samples were loaded on the gel and electrophoresed for 11 h.

The 602 bp region was closed with exactly the same primers on both strands using the standard protocol for the CEQ 2000 sequencer. The CEQ protocol for sequencing was as follows: the DNA sequencing reaction was prepared using the CEQ cycle sequencing chemicals (10× sequencing reaction buffer, dNTP mixture, ddUTP, ddGTP, ddCTP and ddATP dye terminators) and 125 fmol of the DNA sample subcloned into the pGEM vector. The sample solution was then preheated at 76°C for 5 min. The CEQ polymerase (AmpliTaQ) and 3.2 pmol primer were then added to the sample. The solution was first heated at 94°C for 2 min and then amplified at 94°C for 30 s, 55°C for 30 s, 60°C for 4 min, repeated for 34 cycles, warmed to 60°C for 10 min and finally held at 4°C. Stop solution (1.5 M sodium acetate and 50 mM EDTA) and 20 μg/μL of glycogen were added to

Table 1. 32-Base Gap (Highlighted in Bold)

```
TTCTACTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTGAGACGGAGTCTCACTCTGTT
GCCCAGGCTGAAATGCAATGCCACAATCTCAACTCATTGCAACCTCTGCCTCCCGGGTT
GAAGTGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAAGGATGCACCACCAC
ACCCAGCTAATTTTTATATTTTAGTAGAGATGGGGTTTCGCGCACATTGGCCAGGCTG
GTCTTGAACCTCCTGACCTCAGGTGATCCACCTGCCTCAGCCTCCCAAAGTGTGGAATT
ACAGGTGTGAGCCACCGCACCAGGCCTACTCTTTTTAACTCTTTTTTTTTGAGACG
GAGTCTTGCTCTGTGCGCCACTGCACTACAGCCTAGGCGACAGAGCGAGACTCCGTCTCA
AAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACACCCAAGTATCATTT
TACCAAGAGTATCATAATTCAGACTCTCAATGTACCTTGAGGTCTTGAGAATCCGGC
TAAGCAGGCAGTTTTCCAGTTCAGGAGCACCAGAAAGGTAAGGCTTCAGCTTTTT
TCCACAGA
```

the sample before ethanol precipitation. The sample was vacuum dried in a SpeedVac for 10 min and then resuspended in 40 μL of deionized formamide (Sigma, St. Louis, MO, USA). The sample was transferred to the specific polypropylene CEQ sample plate and overlaid with one drop of light mineral oil (Sigma). The plate was placed into the CEQ 2000 sequencer, a new gel syringe (Beckman Coulter) was placed into the gel pump and Method DTCS-3 (180 min per sample plate row) was run.

The completed sequence for the region was assembled using Sequencher (2). Figure 1 shows electropherograms from a subset of the ABI 377 sequence traces and the CEQ 2000 traces in the region of interest. Typical of the ABI 377 analyses was a sudden drop in signal strengths, a hard stop, in both directions (strands) that either resulted in an immediate and complete termination or a gradual decrease of the sequence over approximately 20 bases to noise. Aligned highest quality traces for each approach and the resulting complete analysis through the difficult region are illustrated in the figure.

Inspection of the sequence in Table 1 shows that the 32-base gap (highlighted in bold) not sequenced by the ABI 377 sequencer analyzes in both directions. The next 27 bases were of low peak height but could be determined by the ABI software. The entire 62-base region was flanked by repeats, including a poly T repeat and a CTTT repeat on one side and a CAAA repeat on the other, which give rise to structures difficult for certain combinations of polymerases and dye terminator-labeled

bases to extend. Attempts to use ABI dRhodamine chemistry to resolve AT rich regions (CG/AT = 0.8) did not work. To exhaust the possibilities, we substituted dGTP for dITP in the ABI BigDye chemistry that is typically used to resolve GC rich regions, and obtained the best results (one analysis and only in one direction). This analysis could resolve 20 of the bases, leaving six bases on each end of the region indeterminate or wrong. The CEQ chemistry is based on dITP, not dGTP, and worked well for this region. Denaturing agents, betaine and DMSO, were also used with no effect.

The bases 92 to 312 of the above sequence contain high homology to an Alu repeat of the HS/PV subfamily. These repetitive elements are found throughout the genome with approximately 500 000 copies per haploid genome. Certain elements of the sequence that may contribute to its refractory qualities are also found in Alu elements. It is therefore possible that the sequencing approach described here will be useful in many other refractory sites in the genome. The sequence is from intron 4 of the human homologue (chromosome 17) of the yeast CDC6 gene (8). The CDC6 protein is an important component of the DNA replication apparatus in yeast and appears to contribute to the "licensing" function that allows progression of the cell cycle from G1 to S phase. It is not known if the structural consequences of the unusual inverted repeats seen in this sequence might confer a regulatory role on this region of the gene.

A series of experiments was performed to determine the reasons for the

Benchmarks

Figure 1. Graphic alignment of the electropherograms obtained from the two instruments, the ABI 377 and the CEQ 2000 sequencers. The assembly of the electropherograms is broken in the middle to facilitate reproduction. One CEQ 2000 analysis spans the region. The two ABI 377 analyses selected from all attempts sequence the region on different strands.

improvement in sequencing this difficult region on the CEQ 2000 sequencer. A variety of concentrations of the DNA sample was tried within the recommended protocol range of 50–100 fmol. The best results were obtained when this limit was exceeded and 125 fmol DNA were used. Thermal cycling conditions were surveyed, including preheating (immediate sample denaturation before thermal cycling) conditions. Also changed were the capillary temperatures during the DTCS-3 run method on the CEQ 2000. None of these temperature changes improved the sequencing. Deviation from the recommended temperatures actually tended to cause failures (no resulting sequence) in most cases. We conclude that successful sequencing across the gap resulted from a combination of the cycle sequencing protocol that included preheating, which improves the analysis length and quality of all analyses) and reagents (polymerase enzymes and dyes). While there is no single solution, applying several available DNA sequencing technologies can make finishing DNA sequencing projects more efficient and can also overcome sequence-specific difficulties.

REFERENCES

1. Brown, T.A. 1994. DNA Sequencing: The Basics. IRL, New York.
2. Lisiewicz, J., A. Godany, D.V. Agoston and H. Kuntzel. 1988. Cloning and characterization of the *Saccharomyces cerevisiae* CDC6 gene. *Nucleic Acids Res.* 16:11507-11520.
3. Mytelka, D.S. and M.J. Chamberlin. 1996. Analysis and suppression of DNA polymerase pauses associated with a trinucleotide consensus. *Nucleic Acids Res.* 24:2774-2781.
4. Perkin-Elmer. 1995. ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit Protocol. Appendix C, p. 19. Perkin-Elmer, Eden Prairie.
5. Steffens, D.L. and R. Roy. 1998. Sequence analysis of mitochondrial DNA hypervariable regions using infrared fluorescence detection. *BioTechniques* 24:1044-1046.
6. Taub, E. 1996. Science world reaps rewards with analysis tools for the Macs. *MacWeek* 10:35.
7. Vander Horn, P.B., M.C. Davis, J.J. Cunniff, C. Ruan, B.F. McArdle, S.B. Samols, J. Szasz, G. Hu et al. 1997. Thermo sequenase DNA polymerase and *T. acidophilum* pyrophosphatase: new thermo-stable enzymes for DNA sequencing. *BioTechniques* 22:758-765.
8. Williams, R.S., R.V. Shohet and B. Stillman. 1997. A human protein related to yeast Cdc6p. *Proc. Natl. Acad. Sci. USA* 94:142-147.

Address correspondence to Dr. Harold R. Garner, University of Texas Southwestern Medical Center at Dallas, McDermott

Center for Human Growth and Development and Center for Biomedical Inventions, 5323 Harry Hines Blvd., Dallas, TX 75235-8591, USA. Internet: garner@swmed.edu

Received 4 June 1999; accepted 21 December 1999.

Kari A. Kukanskis, Zakir Siddiquee, Ralph V. Shohet and Harold R. Garner
University of Texas
Southwestern Medical Center
at Dallas
Dallas, TX, USA