Evaluation of DNA Sequencing Ambiguities Using Tetramethylammonium Chloride Hybridization Conditions

BioTechniques 22:1088-1090 (June 1997)

It is often difficult to ascertain the correct DNA sequence in regions prone to GC compression using conventional sequencing approaches. While thermostable DNA polymerases, which permit priming and extension at elevated temperatures, can improve the accuracy within many ambiguous regions, these enzymes do not eliminate electrophoretic mobility artifacts due to GC compression. Sequencing both strands can prove to be informative, yet where two conflicting sequences exist, the possibility that the gene or sequence is polymorphic must be resolved. Most often, a combination of methods must be used to obtain correct DNA sequence data.

We describe a reliable sequence-specific oligonucleotide (SSO)-based tetramethylammonium chloride (TMAC) hybridization assay that allows discrimination of single base-pair differences and is not subject to the electrophoretic artifacts of acrylamide gel-based DNA sequencing. This method, when used in conjunction with conventional DNA sequencing, can accurately determine the correct sequence across problematic areas with greater certainty. In the dot-blot assay described here, the use of TMAC in hybridization eliminates the preferential melting of dA-dT vs. dG-dC base pairs (2). TMAC binds selectively to dA-dT base pairs, altering the dissociation equilibrium and increasing the melting temperature (3). In a solution containing 3.0 M TMAC, this displacement is sufficient to shift the melting temperature of dA-dT base pairs to that of dG-dC base pairs within the SSO (2,8). Thus, it is the length of the SSO and not the sequence that is the crucial parameter in the hybridization (8).

Previously, our group reported the complete genomic sequence of the gene responsible for the predominant form of polycystic kidney disease, *PKD1* (5). The *PKD1* gene maps to

Table 1. Positions and Nature of DNA Sequence Discrepancies

Nucleotide Position ^a	Sequence Discrepancy	SSO Used in the Dot Blot ^b	
50674	CG or GC	EPKDC ^c	AGCA CG TACGCTTCGTG
		APKD1C ^d	AGCA GC TACGCTTCGTG
50819	CG or GC	EPKDC	CGCA CG TGAGCTCCGCA
		APKD1C	CGCA GC TGAGCTCCGCA
51848	CC or CCCC	EPKDC	AGGACA CC —TTCGGGCCA
		APKD1C	GGACA CCCC TTCGGGCC

^aPosition of the discrepancy in the APKD1C genomic sequence.
^bSSO corresponding to the reported sequences. In dot-blot hybridizations, these pairs of oligonucleotides were used as cold competitor SSOs to each other.
^cEPKDC (6).
^dAPKD1C (5).

chromosome 16p13.3. The primary sequence of the interval encompassing the *PKD1* gene is 53 522 bp in length (GenBank® Accession No. 39891). Inspection of our cosmid-derived genomic sequence showed that our sequence differed at three locations from the previously published partial cDNA sequence for the 3' end of the *PKD1* gene (6). Two of the sequence differences cause missense changes, while a third sequence difference causes a frame shift resulting in a new carboxy terminus. One possible explanation for the apparent sequence discrepancies was that the two groups had identified polymorphisms in the *PKD1* gene. An alternative was that the sequence of either group had errors because of technical problems associated with sequencing

G-C-rich segments or artifacts due to cloning or reverse transcription errors. The predicted impact of the protein sequence made it imperative that the issue be definitively resolved.

Following several attempts to resolve these three sequence discrepancies using conventional enzyme-based sequencing, we analyzed the sequences in question with an independent SSO dot-blot assay using TMAC hybridization conditions. We designed sequence-specific 17-mers, complementary to either the published cDNA sequence or our genomic sequence. The SSOs were hybridized to polymerase chain reaction (PCR)-amplified DNA from the genomic clone from which our sequence was derived (cDEB11), an overlapping genomic clone (P1 91.8B),

Figure 1. Dot-blot analysis with SSOs using TMAC hybridization conditions. Individual hybridizations were set up for each 17-mer SSO indicated (rows A–F), targeted to three nucleotide positions at the 3' end of the *PKD1* gene. (A) EPKDC SSO at position 51848; (B) APKD1C SSO at position 51848; (C) EPKDC SSO at position 50674; (D) APKD1C SSO at position 50674; (E) EPKDC SSO at position 50819; (F) APKD1C SSO at position 50819. The DNA targets are listed above each column. Observe that the genomic DNA PCRs only hybridize the SSOs B, D and F. SSOs A, C and E do not hybridize to any of the genomic PCR samples or to the cloned control templates containing the *PKD1* target sequence (P1 91.8B or the cosmid cDEB11). The cosmid cGGG10.2 is a negative control cloned template. SSOs A, C and E only hybridize to the complementary antisense 40-mers spotted on the right side of the dot blot.

a nonoverlapping genomic clone (cGGG10.2) and a series of 9 PCR-amplified genomic DNA samples from unrelated individuals. In addition, individual 40-mer oligonucleotides complementary to each of the SSOs used as probes were spotted onto the membrane as a positive control target for each SSO. All three DNA sequence discrepancies were located in a 1.6-kb segment at the 3' end of the *PKDI* gene (4). The exact positions and nature of the discrepancies are listed in Table 1.

PCR primers (5'-GGTGTGCCT-GCTGCTGTTCG-3' and 5'-ATGCCC-CTGCCTGCTCTCTG-3'), which were specific for the 3' end of the *PKDI* gene, were used to amplify the 1.6-kb genomic fragment. Genomic DNA (500 ng), cosmid DNA (100 pg) or P1 DNA (200 pg) was amplified in a 50- μ L hot-start PCR containing: 1 $\times$ buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 0.001% [wt/vol] gelatin), 500 pM each primer, 200 pM dNTPs and 1.0 mM MgCl₂. Amplification was initiated using the following parameters: 2.5 U of AmpliTaq® DNA Polymerase (Perkin-Elmer, Norwalk, CT, USA) were added after the reaction mixture had been heated to 80°C, followed by 36 cycles at 96°C for 1 min, 64°C for 30 s and 74°C for 2 min.

The dot blots were set up as previously described (4). Five microliters of each PCR mixture or 2.5 ng of each cloned template were mixed with 50 μ L of the denaturation solution; then 50 μ L were spotted into the appropriate well of the dot-blot apparatus onto the nylon membrane (MagnaGraph™; Micron Separations Inc., Westboro, MA, USA). Thirty picomoles of each SSO were radiolabeled with 30 pmol of [γ -³²P]ATP (NEN Life Science Products, Boston, MA, USA) in a 20- μ L kinase reaction (30 min at 37°C) using 10 U of T4 Polynucleotide Kinase (New England Biolabs, Beverly, MA, USA). The reaction was stopped by the addition of 80 μ L of H₂O.

The dot-blotted membranes were prehybridized for 30 min at 52°C in 2.5 mL of TMAC hybridization solution (3.0 M TMAC, 1 mM EDTA, 10 mM NaPO₄, pH 6.8, 5 $\times$ Denhardt's solution, 0.04 mg/mL yeast RNA, 0.1% sodium dodecyl sulfate [SDS]). Radiolabeled SSOs were added to the hy-

bridization at a final concentration of 100 fmol/mL along with a 10-fold excess of cold, unlabeled competitor SSO (see Table 1). The nylon membranes were hybridized for 16 h at 52°C. The membranes were then washed in TMAC wash solution (3.0 M TMAC, 1 mM EDTA, 10 mM NaPO₄, pH 6.8, 0.1% SDS), once at room temperature for 2 min followed by a 60-min wash at 52°C in fresh wash solution. The washed nylon blots were exposed to autoradiographic film for 48 h at -70°C.

The results of the SSO hybridization assay using TMAC conditions are shown in Figure 1. In all cases, the dot-blot analysis using single-base discrimination conditions (4) clearly indicated the correct genomic DNA sequence. As can be seen in the right-hand portion of the figure, each of the radiolabeled SSOs (A-F) only hybridizes to the expected membrane-bound positive control 40-mer and not to any of the other 40-mers. The American Polycystic Kidney Disease Consortium (APKD1C) SSOs (B, D and F) hybridize to the genomic PCR samples and all of the cloned target samples. None of the SSOs complementary to the previously reported sequences (6) (A, C and E) hybridizes to any of the genomic PCR samples or cloned target samples. These data and the strong hybridization to the control complementary 40-mer for each of the SSOs corroborate our genomic sequence. In addition, the hybridization pattern to the genomic samples provides convincing evidence that the sequencing differences we were trying to resolve were not due to common polymorphisms at the specific sites. Subsequent analysis by the European Polycystic Kidney Disease Consortium (EPKDC) has confirmed our genomic sequence (1,7).

This dot-blot assay is a straightforward method for resolving DNA sequence ambiguities and is particularly useful in regions of high G-C content. All of the sequence differences we describe in this report occurred in regions of G-C compression. Sequencing ambiguity in regions of G-C compression is not uncommon. The sequence derived from one strand may be of higher quality than that derived from the complementary strand. Where two conflicting sequences exist generated from differ-

---

Benchmarks

---

ent genomic sources, the possibility that the gene is polymorphic must be determined. We have found that this hybridization approach, originally designed to allow discrimination of single base-pair changes for genetic disease diagnosis, is very effective in resolving DNA sequence ambiguities. In the presence of TMAC, the SSOs are not subject to the mobility artifacts that cause sequence compression, and using arrayed dot-blots, multiple sequences and multiple genomic samples can be analyzed simultaneously, rapidly and inexpensively.

REFERENCES

1. **Harris, P.C., G. Germino, K. Klinger, G. Landes and J. van-Adelsberg.** 1995. The PKD1 gene product [letter]. *Nat. Med.* 1:493.
2. **Melchior, W.B. and P.H. von Hippel.** 1973. Alteration of the relative stability of dA-dT and dG-dC base pairs in DNA. *Proc. Natl. Acad. Sci. USA* 70:298-302.
3. **Shapiro, J.T., B.S. Stannard and G. Felsenfeld.** 1969. The binding of small cations to deoxyribonucleic acid. Nucleotide specificity. *Biochemistry* 8:3233-3241.
4. **Shuber, A.P., J. Skoletsky, R. Stern and B.L. Handelin.** 1993. Efficient 12-mutation testing in the CFTR gene: a general model for complex mutation analysis. *Hum. Mol. Genet.* 2:153-158.
5. **The American PKD1 Consortium (APKD1C).** 1995. Analysis of the genomic sequence for the autosomal dominant polycystic kidney disease (PKD1) gene predicts the presence of a leucine-rich repeat. *Hum. Mol. Genet.* 4:575-582.
6. **The European Polycystic Kidney Disease Consortium (EPKDC).** 1994. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* 77:881-894.
7. **The European Polycystic Kidney Disease Consortium.** 1994. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. Erratum. *Cell* 81:1170.
8. **Wood, W.L., J. Gitschier, L.A. Lasky and R.M. Lawn.** 1985. Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. *Proc. Natl. Acad. Sci. USA* 82:1585-1588.

This work was sponsored in part by a grant from the National Institutes of Health (DK44853 to K.W.K.). We thank Bill Dackowski for the P1 clone, and we are grateful to Steve Kotsopoulos and Cynthia Reeves for helpful discussions. Address correspondence to Timothy Connors, Genzyme Genetics, One Mountain Road, Framingham, MA 01701, USA. Internet: tconnors@world.std.com

Received 11 September 1996; accepted 22 November 1996.

**T.D. Connors, T.C. Burn,
T. VanRaay, G.G. Germino¹,
K.W. Klinger and G.M.
Landes**

*Genzyme Genetics
Framingham, MA*

¹*The Johns Hopkins University
School of Medicine
Baltimore, MD, USA*