Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Heteroduplex formation and S1 digestion for mapping alternative splicing sites

E.N. Ferreira[1], M.C.R. Rangel[2], P.B. Pineda[2], D.O. Vidal[2,3], A.A. Camargo[3], S.J. Souza[3] and D.M. Carraro[2]

[1]Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brasil
[2]Fundação Antônio Prudente, Hospital A.C. Camargo, São Paulo, SP, Brasil
[3]Instituto Ludwig de Pesquisa sobre o Câncer, São Paulo, SP, Brasil

Corresponding author: D.M. Carraro
E-mail: dirce.carraro@hcancer.org.br

Genet. Mol. Res. 7 (3): 958-969 (2008)
Received june 2, 2008
Accepted August 11, 2008
Published September 30, 2008

ABSTRACT. The identification of alternatively spliced transcripts has contributed to a better comprehension of developmental mechanisms, tissue-specific physiological processes and human diseases. Polymerase chain reaction amplification of alternatively spliced variants commonly leads to the formation of heteroduplexes as a result of base pairing involving exons common between the two variants. S1 nuclease cleaves single-stranded loops of heteroduplexes and also nicks the opposite DNA strand. In order to establish a strategy for mapping alternative splice-prone sites in the whole transcriptome, we developed a method combining the formation of heteroduplexes between 2 distinct splicing variants and S1 nuclease digestion. For 20 consensuses identified here using this methodology, 5 revealed a conserved splice site after inspection of the cDNA alignment against the human genome (exact splice sites). For 8 other consensuses, conserved splice sites were mapped at 2 to 30 bp from the border, called proximal splice sites; for the other 7 consensuses, conserved splice sites were mapped at 40 to 800 bp, called distal splice sites. These latter cases showed a nonspecific activity of S1 nuclease in digesting double-strand DNA. From the 20 consensuses identified here,

5 were selected for reverse transcription-polymerase chain reaction validation, confirming the splice sites. These data showed the potential of the strategy in mapping splice sites. However, the lack of specificity of the S1 nuclease enzyme is a significant obstacle that impedes the use of this strategy in large-scale studies.

Key words: Alternative splicing; S1 nuclease; Heteroduplex; Methodology

INTRODUCTION

Alternative splicing is a common feature of human genes, occurring in approximately 60% of them. Its contribution to protein diversity is highlighted by the fact that 80% of the events occur within the coding region, which may affect the function, cellular localization, enzymatic activity, and affinity to substrates of proteins. Many strategies have been used for large-scale characterization of the transcriptional diversity. These strategies are based on cDNA sequence exploration, especially EST databases (Modrek et al., 2001; Xu et al., 2002; Hsu et al., 2005), microarray platforms, either genomic platforms or ones specifically designed for alternative splicing evaluation (Johnson et al., 2003; Gardina et al., 2006; Cuperlovic-Culf et al., 2006; Kwan et al., 2007), and, more recently, construction of cDNA libraries enriched for alternatively spliced exons (Watahiki et al., 2004; Thill et al., 2006). Polymerase chain reaction (PCR)-based approaches are also very useful for alternatively spliced transcript identification (Venables and Burn, 2006; Venables, 2008), especially when used on a small scale. For example, to explore the transcriptional diversity of the VEGF (vascular endothelial growth factor) gene, Eckhart et al. (1999) designed primers at constitutive exons of all variants of this gene, identifying $VEGF_{121}$, $VEGF_{165}$ and $VEGF_{189}$ transcripts (526, 658, and 730 bp, respectively). The formation of a heteroduplex between $VEGF_{121}$ and $VEGF_{165}$ was revealed by the detection of an unexpected band of 600 bp. The heteroduplex was elucidated after experiments of melting and reannealing followed by S1 nuclease treatment.

S1 nuclease is an endo-exonuclease that hydrolyzes both terminal and internal phosphodiester bonds of single-stranded DNA and RNA and has been used for removal of 5' and 3' overhanging single-stranded DNA (Roberts et al., 1979) and hairpin loops from the heteroduplex (Wiegand et al., 1975). Apart from digesting the single-stranded loops, S1 nuclease also nicks the opposite DNA strand. Based on the selective cleavage of single-stranded DNA, this enzyme is largely used for mutation and polymorphism detection from heteroduplexes formed between wild-type and mutant or polymorphic DNAs (Larsen et al., 2001). One of the most used strategies is TILLING (targeting induced local lesions in genomes) (McCallum et al., 2000; Sood et al., 2006; Cooper et al., 2008; Suzuki et al., 2008), in which DNA samples are amplified by PCR and treated with a single-strand specific nuclease that cleaves heteroduplexes. Mutations and/or polymorphisms are detected by electrophoresis. It has been shown that even one single base mutation is sufficient to be recognized by S1 nuclease (Till et al., 2004).

In the present study, in order to establish a strategy for mapping alternative splice-prone sites in a large-scale format, we tested the combination of heteroduplex formation

by splicing variants and S1 endonuclease digestion. We identified 20 sequences with conserved splice sites and 5 of them were selected and validated by reverse transcription (RT)-PCR. In spite of this strategy being efficient in the identification of alternative splice-prone sites, the non-specific cleavage of double-stranded DNA by S1 nuclease limited its employment for large-scale splice site.

MATERIAL AND METHODS

Cell culture

The human breast cell line Hb4aC5.2 used in this study was kindly provided by Professor Dr. Michael O'Hare - Ludwig Institute for Cancer Research - New York. This cell line is derived from normal luminal cells transfected with four copies of the full-length cDNA of the c-erbB-2 gene showing tumor characteristics (Harris et al., 1999). Cells were maintained in RPMI medium supplemented with 100 mL/L fetal bovine serum, 5 µg/mL insulin, 5 µg/mL hydrocortisone and 1 mmol/L L-glutamine in a humidified incubator containing 50 mL/L CO_2 at 37°C. The medium was changed every 2-3 days, and after 10 days total RNA was extracted using CsCl gradient (Glisin et al., 1974). The yield of extracted total RNA was determined with a Kontron 810 spectrophotometer GeneQuant pro (Amersham Pharmacia Biotech), and the integrity was also verified by electrophoresis through 1% agarose gels upon visualization with ethidium bromide.

Double-stranded cDNA synthesis

For double-stranded cDNA (dscDNA) synthesis, the template switch (TS) method was used in order to enrich for full-length cDNA (Matz et al., 1999). TS primer anneals to the cytosine residues inserted by the murine leukemia virus reverse transcriptase at the 3' end of the first strand cDNA, after the enzyme has reached the 5'end of the RNA template. The dscDNA generated by this method contains known sequences at both extremities. For the first strand, 1 µg of total RNA was incubated with 1 µg primer dT(18)-T7 at 70°C for 10 min and snap cooled on ice. The reaction was performed by adding, 1X first strand buffer, 0.01 M dithiothreitol, 40 U of RNasin (Promega, Madison, WI, USA), 1 mM dNTP (GE Healthcare), and 400 U SuperScript II (Invitrogen) in a total volume of 20 µL. The reaction was incubated for 120 min at 42°C. For the second strand synthesis, the Advantage® cDNA PCR Kit (Clontech) was used as follows: 5X cDNA PCR buffer, 1 mM dNTP mix, 5X Advantage cDNA Polymerase Mix, 1.4 U Rnase H (Invitrogen) and 1 µg of template switch primer (5'AAGCAGTGGTAACAACGCAGAGTACGCGGG 3') in a final volume of 100 µL. The reaction was incubated at 37°C for 10 min, 94°C for 3 min, 65°C for 5 min and 75°C for 30 min. The stop reaction solution including 0.25 M NaOH and 0.5 mM EDTA was added, followed by incubation at 65°C for 10 min. The dscDNA was purified by phenol:chloroform:isoamyl alcohol (25:24:1) pH 8.0 extraction.

Polymerase chain reaction

PCR mixture contained 1X buffer, 0.25 mM dNTP, 1.5 mM $MgCl_2$, 1.0 mM of each

primer, 1 M betain and 1 U Platinum® Taq DNA Polymerase (Invitrogen) in a final volume of 20 µL. Thermal cycles and oligonucleotides were specific for each reaction. For the amplification of dscDNA, shorter T7 and TS primers (shT7 and shTS), which anneal to the known extremities (T7 and TS) were used generating dephosphorylated dscDNA. The PCR produt was carried out at 95°C for 4 min followed by: 5 cycles of 95°C for 45 s, 68°C for 1 min and 72°C for 1 min; 5 cycles of 95°C for 45 s, 66°C for 1 min and 72°C for 1 min; 5 cycles of 95°C for 45 s, 64°C for 1 min and 72°C for 1 min, and 15 cycles of 95°C for 45 s, 63°C for 1 min and 72°C for 1 min. The amplification of S1 digested fragments ligated to linker A was performed using three distinct combinations of primers (Figure 2 - item VII): 1) shTS primer + primer A (5' CCATAATGAACAAAGTGAG 3'); 2) shT7 primer + primer A, and 3) primer A alone. The reaction was incubated at 95°C for 4 min followed by 35 cycles of 95°C for 45 s, 60°C for 1 min and 72°C for 1 min, and a final extension at 72° for 7 min. Primers were designed for evaluation of 5 consensuses whose sequences are described in Table 1. The thermal cycle was 95°C for 4 min followed by 35 cycles of 95°C for 45 s, Tm for 1 min and 72°C for 1 min, and a final extension at 72° for 7 min.

Table 1. Oligonucleotides used for reverse transcription-polymerase chain reaction validation.

Gene symbol	Oligonucleotide foward (5'- 3')	Oligonucleotide reverse (5'- 3')	Tm	Amplicon
DGCR6L	CTCAGCAGCGAGAACTAG	GAGTTCCAGCAGGTTCATC	67°C	194 bp
PTMA	CTGCTAACGGGAATGC	CTCGTCGGTCTTCTGC	64°C	224 bp
HAGH	CGAGTACACCATCAACAACCTC	CACCGGGTCCGTCTCAC	71°C	208 bp
FTSJ3	GAGAGAAAGTGGCACAGC	CTTTTGTTCCTTACG	67°C	247 bp
HNRPDL	CGGATATGATTATACTGGG	CAATGTCGTCCTGCAAG	64°C	201 bp

Denaturation and renaturation

dscDNA was heated at 95°C for 20 min and maintained at 42°C for 20 h for total denaturation and renaturation in a buffer containing 0.2% SDS, 0.5 M NaCl, 0.05 M Tris-HCl, pH 7.5, and 30% formamide in a total volume of 30 µL. The sample was purified by GFX PCR and Gel Band Purification Kit (GE Healthcare Life Sciences).

S1 nuclease digestion

The establishment of the optimal concentration of S1 nuclease (Promega) was performed by cleaving a PCR product (dscDNA) of 480 bp with different concentrations of enzyme: 1, 2, 4, 6, 10, 50, 100, and 200 U/µg. Cleavage was performed in 1X buffer [0.05 M sodium acetate (pH 4.5 at 25°C), 0.28 M NaCl, 4.5 mM $ZnSO_4$] and incubated at 37°C for 1 h. Following cleavage, fragments were analyzed on 8% acrylamide gels. For the heteroduplex digestion, 4 U/µg S1 nuclease was used in 1X buffer and incubated at 37°C for 1 h. Following digestion, samples were purified by phenol:chloroform:isoamyl alcohol (25:24:1) pH 8.0 extraction.

Ligation of linker A

Approximately 7.5 µg of linker A was incubated with 1.2 µg dscDNA sample in the

presence of 1X buffer and 4 U T4 DNA ligase (1 U/µL; New England Biolabs) at 16°C for 16 h. For enzyme inactivation, the reaction was incubated at 65°C for 10 min.

Cloning

The PCR products amplified with primer A were cloned with the InsTAclone™ PCR Cloning Kit (Fermentas Life Sciences) following manufacturer recommendations. The plasmids were transformed in DH10β cells (Invitrogen) by electroporation (BioRad). Individual colonies were picked and plasmid DNA was extracted by the alkaline lysis method (Sambrook et al., 1989).

Sequencing

Sequencing reactions were prepared using ABI Prism BigDye Terminator v3.1 Cycle Sequencing Ready Reactions (Applied Biosystems), following the manufacturer recommendations, and run on ABI Prism 3100 (Applied Biosystems).

Bioinformatics analysis

Sequences were aligned against the May 2004 assembly of the human genome sequence available at UCSC Genome Browser (Kent et al., 2002), using the BLAT search tool (http://genome.ucsc.edu/cgi-bin/hgBlat) to verify the presence of conserved splice sites (GT and AG dinucleotides) at the boundary of the cDNA to the linker A. In the case of the absence of conserved splice sites, this alignment was manually compared to the alignment of full-length mRNAs from the RefSeq database (Pruitt et al., 2005) against the human genome, in order to determine the presence of splice sites in the surrounding region. In some cases where more than one RefSeq transcript was described for the same gene, the variant containing the higher number of exons was used. Sequences were also aligned against additional databases of human full-length mRNAs [Genebank (Benson et al., 2004) and Mammalian Gene Collection Full ORF mRNAs (MGC Project Team, Gerhard et al., 2004)], ESTs and predicted transcripts [Ensembl (Hubbard et al., 2002), Geneid (Guigó et al., 1992) and Genscan (Burge and Karlin, 1997; Burge, 1998)], to verify the presence of alternatively spliced transcripts.

RESULTS

In the present study, we tested a strategy for mapping alternative splice sites in the whole transcriptome combining the formation of heteroduplex molecules followed by cleavage of the single-stranded loops by S1 nuclease. In order to acquire specific cleavage of single-stranded loops by S1 nuclease that would assure correct identification of alternative splice sites, we initially established the optimal enzyme concentration that should not cleave dscDNA. We tested different concentrations of enzyme ranging from 1 to 200 U/µg (Figure 1). At enzyme concentrations over 4 U/µg, a slight decrease in the size of the band and even a smear (at higher S1 concentration - 100 and 200 U/µg) is visible on acrylamide gels, which is indicative of nonspecific extremities double-strand cleavage by the enzyme. Therefore, the concentration 4 U/µg was used for further experiments.

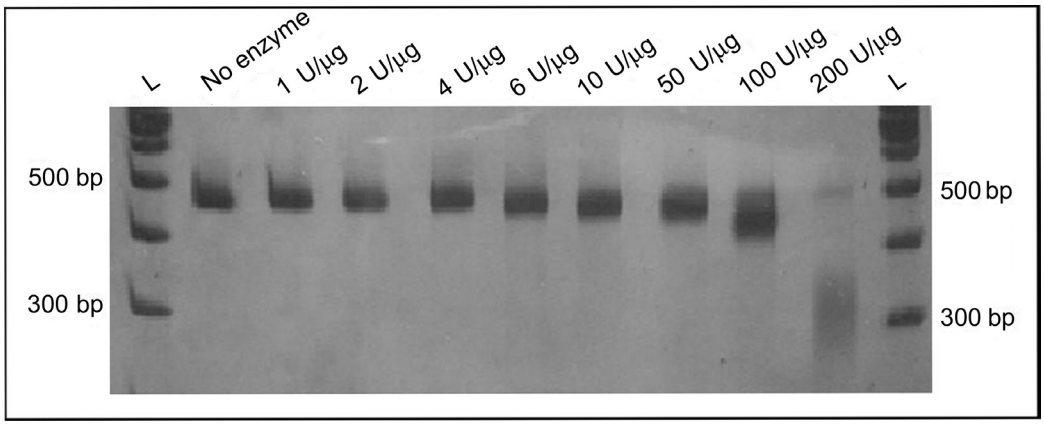

Figure 1. Kinetics of S1 nuclease. A polymerase chain reaction (PCR) product (dsDNA) of 480 bp was subjected to digestion with different concentrations of S1 nuclease for 1 h. As control, a non-digested PCR product is shown. L: 100-bp ladder.

Briefly, total RNA was used for dscDNA synthesis using dT(18)-T7 and TS primer, resulting in dscDNA with known 3' and 5' ends (Figure 2 - I and II). This procedure allowed amplification of dscDNA by the use of complementary primers to the cDNA extremities (shT7 and shTS) (Figure 2 - III). The formation of heteroduplexes was obtained by denaturation and renaturation, permitting the hybridization of complementary regions from two distinct splicing variants from the same gene (Figure 2 - IV). The single-strand loops of the heteroduplexes were then cleaved with S1 nuclease. The cleavage sites of the fragments are indicative of alternative splice sites. S1 nuclease, apart from digesting the single-stranded loops also nicks the opposite DNA strand, resulting in partial fragments of blunt-ended dscDNA showing internal 3'-OH and 5'-PO_4 termini (Figure 2 - V). To the blunt-ended dscDNA, linker A was coupled resulting in fragments containing known sequences at both 5' and 3' ends. Part of molecules containing TS primer at 5' end plus linker A at 3' end, part containing linker A at 5' end plus T7 primer at 3' end, and finally a part containing linker A at both extremities, resulted from molecules that putatively reported 2 events of digestion by S1 (Figure 2 - VI). Since both extremities of all fragments are known, further amplification was feasible using different combinations of primers: 1) shT7 with primer A; 2) shTS with primer A, and 3) only primer A. As PCR positive control, a combination of shTS and shT7 primers was used, which should amplify cDNA molecules that did not form any heteroduplex and consequently was not cleaved by S1 nuclease. As negative control, shT7 or shTS primer was used individually and no amplification was observed (data not shown). A slightly smaller-sized smear was observed, as expected, for the combinations of primers when compared to the positive control (data not shown). In order to test the methodology, only the PCR resulting from primer A alone was used for cloning and sequencing (Figure 2 - VII). In this study, 60 isolated colonies were picked up, and plasmid DNA purified and sequenced. Through a manual analysis, 13 sequences were excluded due to low sequencing quality and presence of empty vectors. A total of 47 high-quality sequences were analyzed identifying linker A in all of them. After trimming linker A, sequences were grouped in 23 consensuses (4 clusters and 19 singlets) by sequence similarity. It was expected that the region encompassing the cDNA fragment/linker

A limit should represent an exon border, confirmed by the presence of conserved splice sites (dinucleotides GT or AG). Therefore, these consensus sequences were aligned against the human genome sequence allowing the definition of exon/intron boundaries. All consensus sequences were aligned against the human genome through BLAT, indicating no artifact generation during the amplification and cloning process, and also against the full-length mRNAs (RefSeq), obtaining the 5'- 3' direction of the consensuses. For 3 of them, no full-length mRNA was identified and they were excluded from further analysis. The gene symbol corresponding to the 20 consensus sequences and RefSeq entries used as reference sequence are described in Table 2. Only in 2 consensuses did both extremities align at internal exons of the full-length mRNAs, suggesting the occurrence of 2 alternative splicing events. In the remaining 18 consensuses (90%), one of the ends aligned at the first or last exon of the full lengths (Table 2), suggesting non-specific dscDNA cleavage of S1 nuclease, which allowed the coupling of linkers at 3' or 5' cDNA extremities. Consequently, these 18 consensuses resulted from only one alternative splicing event.

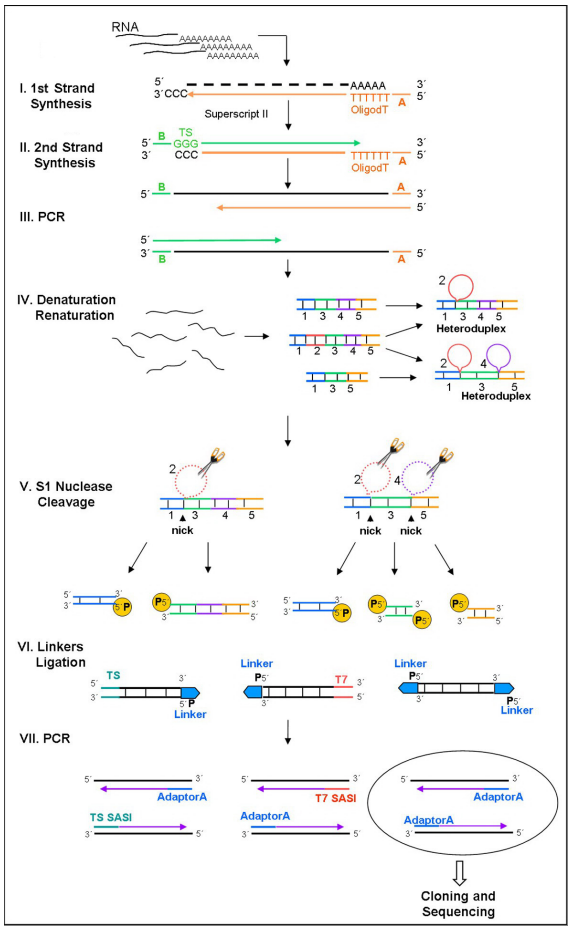

Figure 2. Strategy for the identification of alternative splice sites.

Table 2. Characterization of consensus sequences: alignment to exons of full-length mRNAs and alignment of fragment ends in the exon position of the RefSeq entry. Distance of the conserved splice sites: the position is identified by comparison to full-length mRNAs.

Consensus	Gene symbol	RefSeq	Alignment to exons of full-length mRNAs		Distance from conserved splice sites	RT-PCR confirmed	Presence of alternatively spliced transcripts in the databases
1	DGCR6L	NM_033257.2	5 (3')	4	exact splice site	Yes	No
2	C11orf48	NM_024099.3	7 (3')	6	exact splice site	-	No
3	SNRPD2	NM_177542.1	4 (3')	3	exact splice site	-	No
4	PSME1	NM_006263.2	11 (3')	10	exact splice site	-	No
5	PTMA	NM_001099285.1	5 (3')	4	exact splice site	Yes	Yes
6	C10orf28	NM_014472.4	9 (3')	9	808 bp	-	No
7	RPS25	NM_001028.2	4 (3')	2	18 bp	-	No
8	CLDN7	NM_001307.4	4 (3')	2	97 bp	-	No
9	HAGH	NM_001040427.1	10 (3')	10	20 bp	Yes	No
10	NOLA2	NM_017838.3	4 (3')	2	23 bp	-	No
11	GPSN2	NM_138501.4	9 (3')	6	10 bp	-	Yes
12	PPM1G	NM_002707.3	10 (3')	7	71 bp	-	Yes
13	FTSJ3	NM_017647.2	21 (3')	21	139 bp	Yes	No
14	NDUFS5	NM_004552.1	3 (3')	3	29 bp	-	No
15	RPS11	NM_001015.3	1 (5')	2	44 bp	-	Yes
16	RPL34	NM_033625.2	4	5	Exon 4 - 32 bp Exon 5 - 10 bp	-	Yes
17	HNRPDL	NM_031372.2	9 (3')	9	1 bp Exon 2 - 2 bp	Yes	Yes
18	HINT2	NM_032593.2	2	4	Exon 4 - exact splice site	-	Yes
19	SOX15	NM_006942.1	2 (3')	2	184 bp	-	No
20	PLXNA2	NM_025179.3	32 (3')	32	343 bp	-	No

*The 5' or 3' corresponding to the first or last exon of the full length. (-) Not determined.

Conserved splice sites at the exact border from the alignment of the cDNA against the human genome were only detected for 5 consensuses (denominated exact splice sites). Three of them were also supported by splice sites defining the exon/intron boundaries of full-length transcripts present in the databases. In order to characterize the other 15 consensuses that did not show an exact splice site, we compared the human genomic coordinates against both the consensuses and the corresponding full-length mRNA sequences (RefSeq). For 8 consensuses (53%), the conserved splice sites of full-length mRNAs were observed in a distance from 2 to 30 bp, which was denominated proximal splice sites. For the remaining 7 consensuses (47%), the conserved splice sites of full-length mRNA were present in a distance from 40 to 800 bp, denominated distal splice sites. All the splice sites identified were characterized by the exons defined by them, in the full-length sequence (Table 2). The absence of exact splice sites for the majority of the consensuses is strong evidence that the S1 nuclease nonspecifically cleaves dscDNA in addition to the single-strand loops, even at low enzyme concentration.

From the 20 consensus, 5 were selected for RT-PCR validation, including 2 with exact splice sites (1 supported and 1 not supported by RefSeq data), 2 with proximal splice sites, and 1 with distal splice site (135 bp). The primers were designed at exons corresponding to our consensuses and at subsequent exons based on the full-length mRNA sequences. RT-PCR products were sequenced and all of them confirmed the exon/exon boundaries defined by the splice sites (Table 2). For one of the validated consensuses corresponding to the HNRPDL gene, 2 distinct PCR fragments were amplified

(Figure 3). The 2 alternatively spliced transcripts probably participated in the formation of heteroduplexes. One of the fragments corresponds to an mRNA sequence already identified (NM_031372.2) and the other one is evidence of a novel transcript that defines one additional exon of the gene. This exon adds 105 nucleotides to the mRNA 3'UTR (untranslated region), with no change at the coding sequence. During the course of this study, another alternatively spliced transcript of the HNRPDL gene (NR_003249.1), corresponding to this novel transcript identified in RT-PCR, was deposited at the databases (UCSC Genome Browser - March 2006 assembly), confirming our findings.

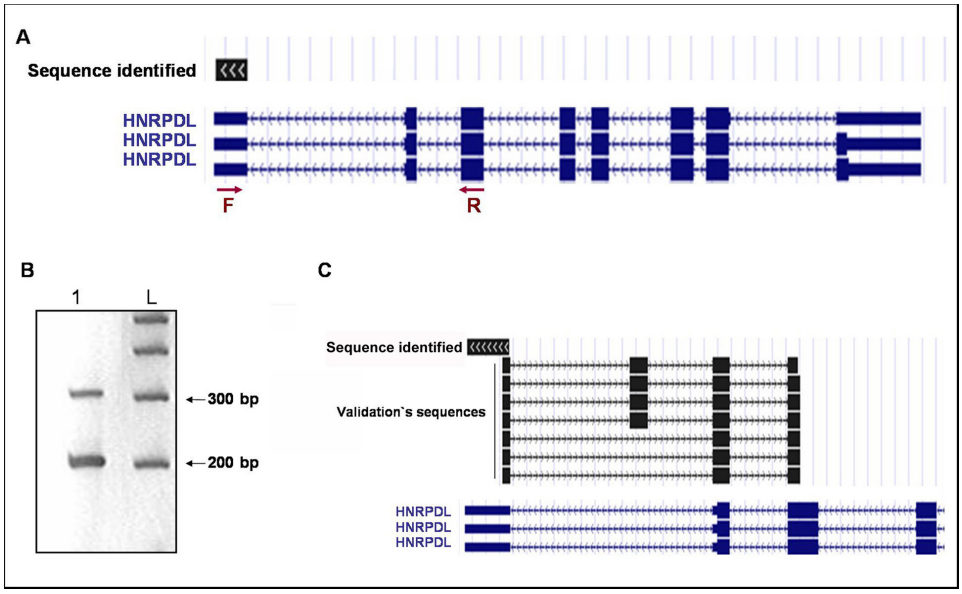

Figure 3. Validation of HNRPDL by reverse transcription-polymerase chain reaction (RT-PCR). **A.** Alignment of the sequence generated by the methodology described against the human genome and UCSC known genes using BLAT tool. The red arrows represent the primers designed for validation (F - forward primer; R - reverse primer). **B.** Lane 1 = 8% acrylamide electrophoresis - PCR amplification of two distinct variants for HNRPDL; L = 100-bp ladder. **C.** Alignment of the sequences generated by RT-PCR against the human genome and UCSC known genes using BLAT tool.

Additionally, we also searched for alternatively spliced transcripts, which could be involved in heteroduplex formation, in the human databases (Genebank, RefSeq, MGC, ESTs, Ensembl, Geneid, and Genescan). For 7 of 20 consensuses (35%), we found 2 alternatively spliced transcripts in the database, which indirectly confirmed the trapping of heteroduplex by our strategy (Table 2). As described in Table 2, two consensuses were validated by both RT-PCR and the presence of alternatively spliced transcripts in databases. No evidence of alternatively spliced transcripts, which supposedly participated in the heteroduplex formation, was described in the databases for the other 13 consensuses (65%), suggesting novel alternative splicing sites.

DISCUSSION

The strategy proposed for alternative splice site mapping was impaired by the occurrence of nonspecific cleavage of S1 nuclease, which was demonstrated by the large number of

proximal and distal splice sites identified. Nevertheless, the power of the strategy was demonstrated by the mapping of 2 novel alternative splice sites, strengthening the occurrence of countless alternatively spliced transcripts not yet identified.

The 5 exact splice sites mapped by this strategy defined alternatively spliced transcripts for DGCR6L (DiGeorge syndrome critical region gene 6-like), C11orf48 (chromosome 11 open reading frame 48), SNRPD2 (small nuclear ribonucleoprotein D2 polypeptide), PSME1 (proteasome activator subunit 1/PA28 alpha) and PTMA (prothymosin, alpha). The splice sites identified for the DGCR6L and PTMA were validated by RT-PCR. Two proximal and one distal splice sites were also confirmed by RT-PCR, strongly suggesting that all 15 nonexact ones represent alternative splice-prone sites. On the other hand, the lack of specific activity of S1 nuclease impeded the use of the strategy for large-scale alternative splice site mapping.

Nonspecific cleavage of S1 is difficult to control. At optimal conditions (pH 4.5 and 100 mM NaCl) for maximal cleavage of single-stranded DNA, double-stranded DNA cleavage has also been reported (Wiegand et al., 1975). S1 activity becomes more stringent at increased salt concentrations (Vogt, 1973). Following this recommendation, S1 digestion was performed in this study under optimal pH (4.5) and high concentration of NaCl (280 mM), in an attempt to control nonspecific cleavage. However, even with this stringent condition, nonspecific cleavage was detected. A further increase in pH and salt concentration would strongly inhibit double-stranded cleavage. Nonetheless, under this condition the rate of activity on single-stranded DNA would drastically decrease (Wiegand et al., 1975), making our strategy unfeasible. Furthermore, at conditions where double-strand cleavage is inhibited, the nick activity of S1 hardly occurs. Yamamoto and Komiyama (2004) also described double-stranded DNA digestion after treatment with S1 nuclease for 1 h or more. In our case, the S1 treatment was performed for 30 min, and even then, the undesired cleavage was not avoided. Therefore, establishing optimal conditions of salt, pH and concentration of enzyme that would specifically cleave single-stranded loops without nibbling double-stranded cDNA termini and also inducing a nick at the opposite strand is a difficult task. Other members of the S1 nuclease family have also been used for mutation detection studies and show more specific action on single-stranded DNA, such as CEL I and mung bean nuclease (Martin et al., 1986). However, they do not show the nick activity that is essential for this study.

It is clear that although promising, the lack of specificity of the S1 nuclease enzyme is a significant obstacle that impedes the use of this strategy for mapping alternative splice sites in large-scale studies. The availability of a nuclease that more specifically cleaves single-stranded cDNA and shows nick activity with no double-strand digestion would make this strategy a fantastic tool for large-scale alternative splice site mapping.

ACKNOWLEDGMENTS

We would like to thank Natanja S. Kirschbaum-Slager for helping in preliminary experiments. Research supported by CEPID-FAPESP (#98/14335-2). E.N. Ferreira and M.C.R. Rangel were recipients of fellowships from FAPESP. D.O. Vidal and P.B. Pineda were recipients of fellowships from CNPq and CAPES, respectively.

REFERENCES

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. (2004). GenBank: update. *Nucleic Acids Res.* 32: D23-D26.
Burge CB (1998). Modeling Dependencies in Pre-mRNA Splicing Signals. In: Computational Methods in Molecular

Biology (Salzberg S, Searls D and Kasif S, eds.). Elsevier Science, Amsterdam, 127-163.
Burge C and Karlin S (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268: 78-94.
Cooper JL, Till BJ, Laport RG, Darlow MC, et al. (2008). TILLING to detect induced mutations in soybean. *BMC Plant Biol.* 24: 8-9.
Cuperlovic-Culf M, Belacel N, Culf AS and Ouellette RJ (2006). Microarray analysis of alternative splicing. *OMICS* 10: 344-357.
Eckhart L, Ban J, Ballaun C, Weninger W, et al. (1999). Reverse transcription-polymerase chain reaction products of alternatively spliced mRNAs form DNA heteroduplexes and heteroduplex complexes. *J. Biol Chem.* 274: 2613-2615.
Gardina PJ, Clark TA, Shimada B, Staples MK, et al. (2006). Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* 7: 325.
Gerhard DS, Wagner L, Feingold EA, Shenmen CM, et al. (2004). The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res.* 14: 2121-2127.
Glisin V, Crkvenjakov R and Byus C (1974). Ribonucleic acid isolated by cesium chloride centrifugation. *Biochemistry* 13: 2633-2637.
Guigó R, Knudsen S, Drake N and Smith T (1992). Prediction of gene structure. *J. Mol. Biol.* 226: 141-157.
Harris RA, Eichholtz TJ, Hiles ID, Page MJ, et al. (1999). New model of ErbB-2 over-expression in human mammary luminal epithelial cells. *Int. J. Cancer* 80: 477-484.
Hsu FR, Chang HY, Lin YL, Tsai YT, et al. (2005). AVATAR: a database for genome-wide alternative splicing event detection using large scale ESTs and mRNAs. *Bioinformation* 1: 16-18.
Hubbard T, Barker D, Birney E, Cameron G, et al. (2002). The Ensembl genome database project. *Nucleic Acids Res.* 30: 38-41.
Johnson JM, Castle J, Garrett-Engele P, Kan Z, et al. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302: 2141-2144.
Kent WJ, Sugnet CW, Furey TS, Roskin KM, et al. (2002). The human genome browser at UCSC. *Genome Res.* 12: 996-1006.
Kwan T, Benovoy D, Dias C, Gurd S, et al. (2007). Heritability of alternative splicing in the human genome. *Genome Res.* 17: 1210-1218.
Larsen LA, Andersen PS, Kanters J, Svendsen IH, et al. (2001). Screening for mutations and polymorphisms in the genes KCNH2 and KCNE2 encoding the cardiac HERG/MiRP1 ion channel: implications for acquired and congenital long Q-T syndrome. *Clin. Chem.* 47: 1390-1395.
Martin SA, Ullrich RC and Meyer WL (1986). A comparative study of nucleases exhibiting preference for single-stranded nucleic acid. *Biochim. Biophys. Acta* 867: 76-80.
Matz M, Shagin D, Bogdanova E, Britanova O, et al. (1999). Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res.* 27: 1558-1560.
McCallum CM, Comai L, Greene EA and Henikoff S (2000). Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123: 439-442.
Modrek B, Resch A, Grasso C and Lee C (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.* 29: 2850-2859.
Pruitt KD, Tatusova T and Maglott DR (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 33: D501-D504.
Roberts TM, Kacich R and Ptashne M (1979). A general method for maximizing the expression of a cloned gene. *Proc. Natl. Acad. Sci. U. S. A.* 76: 760-764.
Sambrook J, Fritsch E and Maniatis T (1989). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.
Sood R, English MA, Jones M, Mullikin J, et al. (2006). Methods for reverse genetic screening in zebrafish by resequencing and TILLING. *Methods* 39: 220-227.
Suzuki T, Eiguchi M, Kumamaru T, Satoh H, et al. (2008). MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. *Mol. Genet. Genomics* 279: 213-223.
Thill G, Castelli V, Pallud S, Salanoubat M, et al. (2006). ASEtrap: a biological method for speeding up the exploration of spliceomes. *Genome Res.* 16: 776-786.
Till BJ, Burtner C, Comai L and Henikoff S (2004). Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Res.* 32: 2632-2641.
Venables JP (2008). Enrichment of alternatively spliced isoforms. *Methods Mol. Biol.* 419: 161-170.
Venables JP and Burn J (2006). EASI - enrichment of alternatively spliced isoforms. *Nucleic Acids Res.* 34: e103.
Vogt VM (1973). Purification and further properties of single-strand-specific nuclease from *Aspergillus oryzae*. *Eur. J.*

Biochem. 33: 192-200.

Watahiki A, Waki K, Hayatsu N, Shiraki T, et al. (2004). Libraries enriched for alternatively spliced exons reveal splicing patterns in melanocytes and melanomas. *Nat. Methods* 1: 233-239.

Wiegand RC, Godson GN and Radding CM (1975). Specificity of the S1 nuclease from *Aspergillus oryzae*. *J. Biol. Chem.* 250: 8848-8855.

Xu Q, Modrek B and Lee C (2002). Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* 30: 3754-3766.

Yamamoto Y and Komiyama M (2004). Combination of S1 nuclease and PNA for site-selective hydrolysis of double-stranded DNA. Comparison with the site-selective hydrolysis using Ce(IV)/EDTA. *Nucleic Acids Symp. Ser.* 48: 149-150.