Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 8995–8999, November 1990
Genetics

Exon trapping: A genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA

(retroviral vectors/RNA splicing/human genetics)

GEOFFREY M. DUYK*†, SUWON KIM‡, RICHARD M. MYERS§¶, AND DAVID R. COX*‡¶

Departments of *Pediatrics, ‡Psychiatry, §Physiology, and ¶Biochemistry/Biophysics, The University of California at San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143-0554

Communicated by Thomas P. Maniatis, July 31, 1990

ABSTRACT Identification and recovery of transcribed sequences from cloned mammalian genomic DNA remains an important problem in isolating genes on the basis of their chromosomal location. We have developed a strategy that facilitates the recovery of exons from random pieces of cloned genomic DNA. The basis of this "exon trapping" strategy is that, during a retroviral life cycle, genomic sequences of nonviral origin are correctly spliced and may be recovered as a cDNA copy of the introduced segment. By using this genetic assay for cis-acting sequences required for RNA splicing, we have screened ≈20 kilobase pairs of cloned genomic DNA and have recovered all four predicted exons.

Classic genetic analysis of mouse and human has identified many mutant loci for which the biochemical basis of the given phenotype has not been defined. Construction of detailed genetic and physical maps delineate the minimal region within which a gene of interest may be found. The flanking markers demarcating the boundaries of this region are often separated by several hundred kilobase pairs, and candidate genes are selected on the basis of their location between these markers. This approach has been successful in the identification of a number of genes, including those for Duchenne muscular dystrophy (1) and cystic fibrosis (2).

In spite of these successes, the identification and recovery of transcribed sequences remain important technical problems. Experience gained in the cloning of the genes noted above and efforts to recover all transcription units in the mouse and human major histocompatibility complex (3-5) have established three screening strategies for genes from cloned genomic DNA. In vertebrates, most constitutively expressed genes and some regulated coding sequences are marked at their 5′ ends by distinctive regions containing a high density of hypomethylated CpG residues (6). These "CpG islands" are identified by the clustering of recognition sites for rare-cutting restriction enzymes and are confirmed by testing with methylation-sensitive and -insensitive isoschizomers (6). A second method for identifying coding sequences is interspecies cross-hybridization ("zooblots"; ref. 1). Many, but not all, unique and low-copy sequences conserved between species represent genes and can be detected by Southern hybridization techniques. A third method is direct screening of cDNA libraries or Northern blots with whole phage or cosmid clones. The relative merits of these three approaches have been reviewed (7). Other strategies, including DNA sequencing and genetic screens for open reading frames (8), enhancers (9), or promoters (10, 11) are available, but their value appears limited for identifying genes in long segments of cloned genomic DNA.

Mapping and recovering the complete set of transcribed sequences, corresponding to the predicted 2–4% of the genome representing coding sequence, will require additional methods. In this paper, we describe a strategy, exon trapping, that facilitates the recovery of transcribed sequences in cloned mammalian genomic DNA through the functional identification of cis-acting sequences required for RNA splicing.

MATERIALS AND METHODS

Bacteria and Cell Lines. *Escherichia coli* DH5α was used in all instances unless otherwise noted (12). *E. coli* GM48, a Dam methylase-deficient strain (12), was used to prepare pETV-SD DNA. High-efficiency competent bacterial cells were prepared by the method of Hanahan with the modifications of Sambrook *et al.* (12). Cell lines ψ-2, PA-317, and COS were grown and maintained as described (13–15).

Plasmids. pETV-SD is described in Fig. 1*A*. pETV-SD:HBG(+) contains exon 2 of the HBG gene (nucleotide positions 192–477; ref. 17) cloned into the *Bcl* I site of pETV-SD in the sense orientation. pETV-SD:HBG(−) is a similar plasmid with the entire exon–intron–exon motif in the opposite (antisense) orientation. pETV-SD:RGRex.5 plasmids contain a 1.75-kb *Bam*HI–*Bgl* II fragment containing exon 5 of the rat glucocorticoid receptor gene (18) cloned into the *Bcl* I site of pETV-SD in the sense (+) or antisense (−) orientation. pETV:HLA1.5 plasmids contain a 1.527-kb *Bst*YI fragment containing exons 4–6 of the HLA-A2 gene (19) cloned into the *Bcl* I site of pETV-SD in the sense (+) or antisense (−) orientation.

For construction of the pETV-SD:HLA library and subclones, the plasmid pHLA-A2 (19), a 5.1-kb genomic subclone in pUC9 containing a complete copy of the HLA-A2 allele, was digested to completion with *Bst*YI and ligated to *Bcl* I-digested pETV-SD. The ligation mixture was digested with *Bcl* I and transformed into competent *E. coli* DH5α. This step significantly reduces the background of colonies with no inserts, since vector ligated to itself is sensitive to cleavage by *Bcl* I. Ligation of insert to vector destroys the *Bcl* I site, so that recombinant molecules are not linearized. In addition, cleavage of internal *Bcl* I sites in the insert is prevented by *in vivo* Dam methylation of cleavage sites. Analysis of 100 colonies demonstrated that the library contained all expected DNA fragments.

Exon Trapping. Each confluent culture (100-mm plate) of ψ-2 cells was split 1:5 and the next day the cells were transfected with 20 μg of plasmid DNA (pETV-SD or one of its derivatives) in the presence of Lipofectin reagent (ref. 20; BRL). Three days later, cells were split 1:10 into medium

Abbreviations: HBG, human β-globin; X-Gal, 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside; IPTG, isopropyl β-D-thiogalactopyranoside; SA, splice acceptor; SD, splice donor; IVS, intervening sequence; α-β-Gal, α-complementing factor of *E. coli* β-galactosidase; SV40, simian virus 40; PCR, polymerase chain reaction.

†To whom reprint requests should be addressed at: HSE 1556, Box 0554, University of California, 513 Parnassus Avenue, San Francisco, CA 94143.

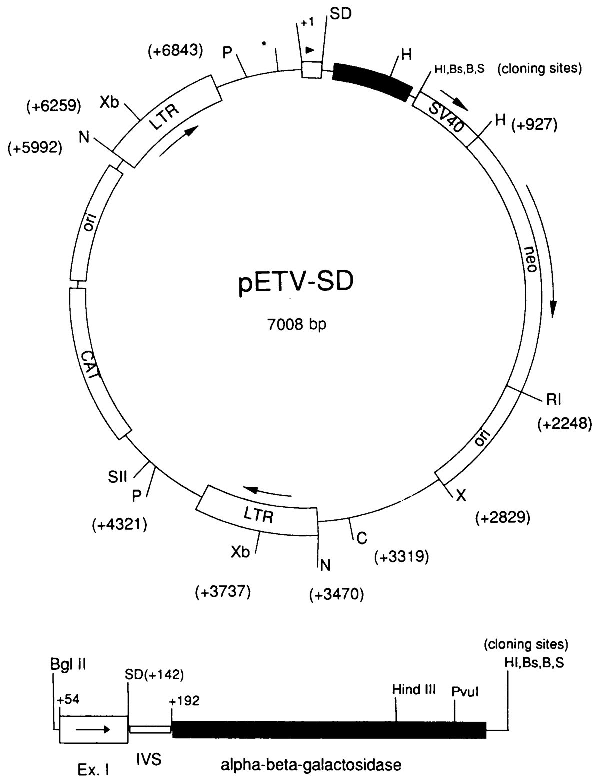

FIG. 1. Maps of pETV-SD and the exon-trap cassette. (*Upper*) Plasmid pETV-SD is a derivative of pDOL⁻ (16), a retroviral shuttle vector with deletion of the naturally occurring splice donor (SD) and splice acceptor (SA) sites. The polyoma origin of replication and early region were replaced with a 1.67-kilobase pair (kb) *Nhe* I–*Pvu* I fragment containing a ColEI origin of replication and a chloramphenicol acetyltransferase (CAT) gene. A 583-base-pair (bp) *Bgl* II–*Sal* I fragment containing the exon-trap cassette was inserted between unique *Bam*HI and *Sal* I sites, resulting in the 7-kb vector pETV-SD. LTR, long terminal repeat; star, retroviral packaging site; open box from +1 to SD, human β-globin (HBG) exon 1 (from +54 to +142); black box, gene encoding the α-complementing factor of *E. coli* β-galactosidase (α-β-GAL); SV40, simian virus 40 origin of replication/early promoter; neo, aminoglycoside phosphoribosyltransferase gene from *E. coli* Tn*5*; ori, ColEI origin of replication; CAT, CAT gene of *E. coli* Tn*9*. The first nucleotide of the exon-trap cassette is defined as position +1 in pETV-SD. Arrows indicate direction of transcription. H, *Hin*dIII; HI, *Bam*HI; B, *Bcl* I; Bs, *Bst*XI; S, *Sal*I; RI, *Eco*RI; X, *Xho* I; C, *Cla* I; N, *Nhe* I; Xb, *Xba* I; P, *Pvu* I; SII, *Sac* II. (*Lower*) Exon-trap cassette. Starting with the 5′ *Bgl* II site, the cassette (583 bp) consists of positions +54 to +192 of the HBG gene in the sense orientation, inserted at an *Eco*RV site of a polylinker, and a 404-bp fragment encoding the α-β-Gal gene followed sequentially by *Bam*HI, *Bst* XI, *Bcl* I, and *Sal* I sites. The *Bgl* II site was destroyed during insertion of the exon-trap cassette into the vector. The cassette contains the wild-type HBG SD with exon and intron sequences previously demonstrated to be required for efficient splicing (17). The SA-site complex for HBG exon II has been deleted and replaced with the α-β-Gal gene. Ex. I, exon 1 of the HBG gene from positions +54 to +142; IVS, the HBG intervening sequence 1 from positions +143 to +192. The *Bam*HI, *Bst*XI, *Bcl* I, and *Sal* I sites in the cassette are unique in pETV-SD.

supplemented with G418 (GIBCO) at 1 mg/ml. After 3–5 days, the G418 concentration was halved and cells were grown as a mixed population for 7 days. Approximately 500 foci per plate were typically observed at this time. This mixed population of cells was grown for an additional 3 days in the absence of G418, the medium was filtered (0.45-μm filter; Nalge Co., New York), and Polybrene (8 μg/ml; Sigma) was added to the filtrate.

PA-317 cells at 20% confluency were infected with 10 ml of the virus-containing filtrate. Four hours later, the filtrate was replaced with medium and the infection protocol was repeated the next morning. After this second round of infection, G418 was used to select for virus-producing colonies.

The virus-containing medium from these colonies was used to infect COS cells by using the infection protocol described above. Forty-eight hours after infection of COS cells, amplified episomal DNA was isolated by the procedure of Hirt (21). The recovered episomal DNA was digested with *Dpn* I, which cleaves any contaminating Dam-methylated plasmid DNA. Episomes replicated in COS cells are insensitive to cleavage as they are not methylated at Dam sites. The digested mixture was used to transform *E. coli* DH5α cells and transformants were selected on LB agar plates containing kanamycin (50 μg/ml), 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal, 0.002%), and isopropyl β-D-thiogalactopyranoside (IPTG, 0.2 mM).

Lac⁻ transformants (white colonies) were patched onto LB/kanamycin/X-Gal/IPTG plates, transferred to nitrocellulose filters, and hybridized to an HBG exon 1-specific probe. Colonies that hybridized to the probe contained plasmids that potentially underwent splicing events and were subsequently analyzed by DNA sequencing. Those that did not hybridize to the probe had undergone gross rearrangement and were not characterized further. DNA was sequenced by the dideoxynucleotide chain-termination method with the Sequenase version 2.0 kit (United States Biochemical). DNA sequencing was primed by using the HBG exon I-specific oligodeoxynucleotide HBG I (5′-GGAGAAGTCTGCCGTTACTG-3′).

Polymerase Chain Reaction (PCR). A PCR assay for delineating spliced versus unspliced products of recovered pETV-SD:HBG(+/−) was developed. Reaction conditions were as described (22). Reactions were run for 30 cycles with the following parameters: denaturation at 93°C, 1 min; annealing at 56°C, 1 min; extension at 70°C, 3 min. Primers were HBG I and HBG II (5′-CCTTCACCTTAGGGTTGCCC-3′), and the sizes of the PCR products were determined by agarose gel electrophoresis. A PCR with a recovered clone containing a "spliced" insert as template results in a 165-bp product and an "unspliced" clone results in a 700-bp product.

RESULTS

Experimental Strategy. pETV-SD is an exon-trap vector that identifies functional SA sites encoded in cloned genomic DNA fragments. Since most genes undergo RNA splicing, such sites serve as identifiers for the majority of genes.

(*i*) Genomic DNA to be tested is "shotgun-cloned" into the vector pETV-SD downstream from the exon trap (Fig. 1). The exon trap consists of a functional SD from the HBG gene and an IVS incorporating the α-β-Gal gene followed by a multiple cloning site (Fig. 1). The vector also contains the cis-essential components for retroviral replication, the SV40 and ColEI origins of replication, and the Tn*5* *neo* gene. The latter confers kanamycin resistance in bacterial cells and G418 resistance in animal cells.

(*ii*) Pooled plasmid DNA from this shotgun cloning is transfected into an ecotropic retroviral packaging cell line, ψ-2 (13). Retroviral packaging cell lines provide the protein products required for propagation of the vector as a retrovirus. The retroviral DNA is transcribed *in vivo* and transcripts derived from recombinant molecules that contain a functional SA may undergo a splicing event with loss of the marked IVS (Fig. 2).

(*iii*) Both spliced and unspliced viral RNAs are packaged into virions, harvested from the medium, and used to infect the amphotropic retroviral packaging cell line PA-317 (14). This step results in an additional round of retroviral replication and produces viral stocks of increased titer capable of infecting COS cells (14, 15, 21). Splicing of cloned genomic

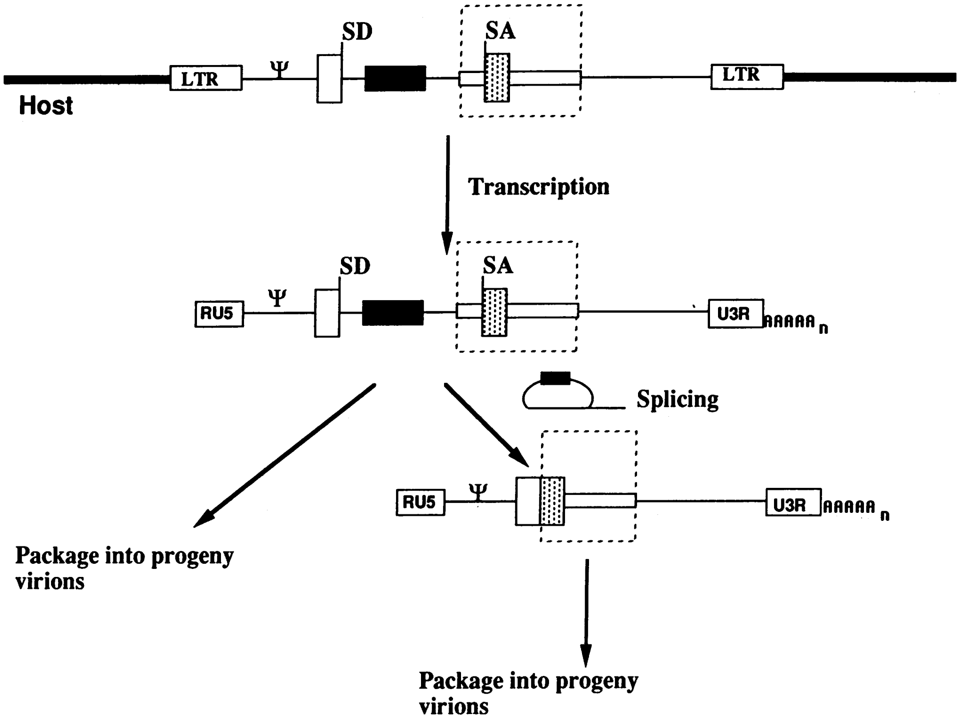

FIG. 2. Life cycle of pETV-SD. A functional SA (enclosed in dashed box) has been cloned adjacent to the exon-trap cassette (vertical open box and horizontal black box). The primary transcription product of the integrated provirus may be packaged directly into progeny virions or may be spliced, with concomitant loss of the marked IVS, and then packaged into progeny virions. When pETV-SD contains a cloned insert without a functional SA, the primary transcript is packaged into the virion without loss of the marked IVS. LTR, long terminal repeat; Ψ, retrovirus packaging site; vertical open box, HBG exon 1; SD, SD of HBG exon I; horizontal black box: α-β-Gal gene; dashed box, randomly inserted fragment containing a functional SA: stippled box, candidate exon; RU5, 5′ leader of retrovirus transcript; U3R, 3′ segment of retrovirus mRNA: $AAAAA_n$, poly A tail.

inserts in retroviral vectors is inefficient and this second round of replication increases the likelihood that a splicing event will occur (21).

(*iv*) Virus isolated from this second cell line is used to infect COS cells, which constitutively produce SV40 large tumor (T) antigen (15). The viral RNA genome is reverse-transcribed and is then amplified as a circular DNA episome due to the presence of the SV40 origin of replication in the vector (Fig. 1).

(*v*) The replicated episomal DNA is recovered from the COS cells, digested with *Dpn* I, and transformed into bacterial cells. Transformants are selected on plates containing kanamycin and X-Gal, a chromogenic substrate for β-galactosidase. Hydrolysis of X-Gal by β-galactosidase produces the characteristic blue color indicative of a Lac^+ phenotype, whereas colonies that do not contain functional β-galactosidase are white.

(*vi*) Only white colonies (Kan^R/Lac^-) are studied further. White colonies result either from inactivation of the α-β-Gal gene by a mutational event or from the loss of the gene as a consequence of a splicing event during the RNA phase of the retroviral life cycle (Fig. 2).

(*vii*) Most mutations are deletions of the exon-trap portion of the vector (see below) and are rapidly detected by colony hybridization with an HBG probe (donor exon).

(*viii*) Bona fide splicing events are verified by direct DNA sequencing primed from within the exon of the splice donor. Correct splicing is indicated by precise removal of the genetically marked IVS and joining of the HBG exon to the "trapped" exon. This candidate exon is an identifier for a potential gene.

Initial Tests with Vector Alone. In initial experiments, we tested whether pETV-SD could be transmitted as a retrovirus in animal cells and recovered as a plasmid in *E. coli*. Following transfection of ψ-2 cells with pETV-SD or its derivatives and infection of PA-317 cells, populations of virus-producing cells regularly yielded titers of 10^2–10^4 colony-forming units/ml. Following infection of COS cells with this virus, episomal DNA from a 100-mm plate of COS cells typically yielded 100–1000 bacterial colonies.

Of the 84 colonies examined, 73 were blue (Kan^R/Lac^+; Table 1). However, a significant number (11/84) were white (Kan^R/Lac^-; Table 1). These white colonies could have arisen by functional inactivation of the α-β-Gal gene by mutation or by loss of function due to cryptic splicing.

Table 1. Summary of results obtained in exon trapping experiments

Clones	No. white/total	No. retaining exon 1/no. tested	No. with spliced products/no. sequenced
Vector	11/84 (13.1%)	2/11 (18.2%)	0/2 (0%)
HBG(+)	385/2030 (19%)	73/82 (89%)	72/73 (98.6%)
HBG(−)	47/504 (9.3%)	7/47 (14.9%)	0/7 (0%)
RGR5(+)	17/63 (27%)	13/17 (76.5%)	9/13 (69.2%)
RGR5(−)	19/107 (17.8%)	3/19 (15.8%)	0/3 (0%)
HLA1.5(+)	26/47 (55.3%)	23/26 (88.5%)	23/23 (100%)
HLA1.5(−)	75/625 (12%)	1/75 (1.3%)	0/1 (0%)
HLA library	345/2686 (12.8%)	86/345 (24.9%)	69/86 (80.2%)

Column 1: "Vector" refers to pETV-SD and the remainder of the column refers to inserts cloned into this vector. *Column 2*: Ratio of recovered Kan^R/Lac^- (white) colonies to total number of Kan^R (white plus blue) colonies. *Column 3:* Ratio of Kan^R/Lac^- (white) colonies that hybridize to the HBG exon 1 probe to the total number of Kan^R/Lac^- (white) colonies tested. *Column 4:* The numerator represents the number of Kan^R/Lac^- (white) colonies that have undergone a splicing reaction resulting in the loss of the marked IVS and ligation of the HBG exon 1 SD to a novel SA. The denominator represents the total number of Kan^R/Lac^- (white) colonies maintaining exon 1 of HBG that were tested.

Cryptic splicing is the utilization of RNA sequences as splice sites that are not normally used in correct processing of wild-type pre-mRNA (23). Restriction mapping of plasmid DNA recovered from these white colonies indicated that the majority (9/11) were grossly rearranged (Table 1). Southern hybridization indicated that the rearrangements deleted the entire cassette. The 2 white colonies that were not grossly rearranged were analyzed by DNA sequencing. In both cases, the HBG SD and exon/intron boundary were intact. Thus, cryptic splicing had not occurred and inactivation of the α-β-Gal gene was due to a point mutation or small rearrangement not detected by Southern blotting or restriction mapping. This relatively high frequency of both gross rearrangements and apparent point mutation is consistent with previous reports (24–26).

Tests with Defined Exons. A set of experiments was performed to determine whether the exon trapping strategy could detect exons in segments of well-characterized cloned genes. Three derivatives of pETV-SD were constructed: pETV-SD:HBG(+), pETV-SD:HLA1.5(+), and pETV-SD:RGR(+). Analogous plasmids with the same DNA fragments in the antisense orientation relative to the exon trap cassette were also constructed. These plasmid DNA molecules were individually processed and analyzed as described above. In view of the observation that the predominant class of mutation in the previous experiment resulted in the loss of the exon-trap cassette by gross rearrangement, we first analyzed white colonies obtained in these experiments by hybridization to a donor exon probe. White colonies that hybridized to the probe were analyzed further by DNA sequencing to determine whether they resulted from genuine splicing events or were the consequence of a mutational event.

In each of the sense-orientation pETV-SD derivatives, recovered exons were correctly spliced (Table 1). For example, when pETV-SD:HBG(+) was used, 2030 colonies were obtained and 385 (19%) were white. We further analyzed 82 of these white colonies by the hybridization screen, and 73 (89%) hybridized to a probe that detects the donor exon in the exon-trap cassette. DNA sequence analysis of 10 of these 73 colonies indicated that all 10 were the result of a precise splicing event. A rapid PCR assay was used to analyze the remainder of these white colonies, and the combined data indicated that 72/73 (98.6%) were the result of bona fide splicing events (Table 1). The plasmid DNA not representing a splice event had an intact exon/intron boundary in the exon-trap cassette, indicating that it did not arise as a consequence of cryptic splicing.

pETV-SD derivatives containing HBG, HLA, and glucocorticoid receptor exons in the antisense orientation relative to the exon-trap cassette were also tested. In all cases, white colonies arose as a consequence of mutation rather than cryptic splicing events (Table 1). Furthermore, almost all (130/141) of these mutation events were gross rearrangements and were detected by colony hybridization. The remaining 11 (7.8%) white colonies were sequenced and found to have an intact exon/intron boundary in the exon-trap cassette, demonstrating that they arose as a consequence of mutation rather than cryptic splicing.

Exon Trapping from a Mixture of Cloned DNA Fragments. We designed a reconstruction experiment to determine whether or not the exon trapping strategy could be used to retrieve candidate exons from a mixture of cloned genomic DNA fragments. A 7.8-kb pUC 9-derived plasmid containing a 5.1-kb cloned genomic DNA fragment of the HLA-A2 gene was shotgun-cloned into pETV-SD (19). The resulting plasmid library contained at least nine different DNA fragments cloned in both orientations. On the basis of experimental design, only 2 of the 18 different classes of plasmid-derived inserts should be recovered as spliced products during the exon trapping procedure. Plasmid DNA from the pooled library was processed as described above. White colonies numbering 345 out of 2686 total colonies (12.8%) were identified and screened for gross rearrangements by colony hybridization (Table 1). The donor exon was maintained by 86 (24.9%) of these colonies and these were analyzed further. A single "G reaction" of DNA sequencing was performed on plasmid DNA from these colonies, which allowed the sorting into groups based upon the pattern of guanine residues. Plasmids recovered from white colonies that yield a G sequence pattern characteristic of the donor exon/intron boundary are the products of a mutation rather than a splicing event. In contrast, a G pattern showing precise loss of the IVS from the donor exon in the exon-trap cassette is diagnostic of a genuine splicing event. G-reaction DNA sequencing identified five groups of plasmids. The exon/intron boundary was retained by 17/86 (19.8%) of these white colonies and thus they arose as a consequence of mutation rather than splicing. The remaining four groups of plasmids represent apparent splicing events (Table 2) and the complete DNA sequence from a single member from each of these groups was obtained. Plasmids in groups A (24/69) and B (5/69), respectively, contain *HLA-2A* exons III and IV. The retrieval of these exons was predicted by the experimental design. However, plasmids in groups C (21/69) and D (19/69) contain inserts that arose by cryptic splicing events. Group C plasmids contain the donor exon of the exon-trap cassette adjacent to a cryptic SA from the antisense orientation of exon III from *HLA-2A* (Table 2). The SA in plasmids from group D is from the β-lactamase gene of pUC9 (27).

Table 2. Recovered splice junctions from HLA library

Class	Sequence	
pETV-SD	SD (142) · GGCCCTGGGAG	gttggtatca
A (HLA exon III)	GGCCCTGGGAG	· SA (1239) GTTCTCACAC
B (HLA exon IV)	GGCCCTGGGAG	· SA (2114) ACGCCCCAAA
C [HLA exon III (−)]	GGCCCTGGGAG	· SA (1480) GTATCTGCGG
D (pUC9)	GGCCCTGGGAG	· SA (1688) TTGCCTGACT

Uppercase letters indicate exon sequences and lowercase letters indicate intron sequences. Vertical line defines exon/intron or exon/exon boundaries. Predicted SD and SA sites are based upon comparison of DNA sequences of cDNA and genomic clones (17, 19, 27) or inferred from this study. The numbers correspond to nucleotide positions and are based upon published sequence data (17, 19, 27). pETV-SD represents the exon/intron sequence at its SD site in the absence of a splicing event. This sequence corresponds to the wild-type exon 1/intron 1 boundary of the HBG gene (17). Groups A–D correspond to the DNA sequences at the junction between the donor exon derived from pETV-SD and the trapped exons. In all cases, the predicted HBG SD was used. Groups A and B correspond to HLA exons III and IV and, in each case, the established SA sites were used. Groups C and D correspond to cryptic SA sites. Group C is found within the predicted antisense orientation of HLA exon III, and group D is found in pUC9.

DISCUSSION

We have used exon trapping to screen ≈20 kb for the presence of known SA sites. The screen identified all four of the predicted wild-type SA sites, whereas only two cryptic splicing events were identified. Consistent with previous reports (21, 24, 26, 28), splicing is not 100% efficient in this retrovirus-based system; however, it occurs frequently enough so that it is readily detected (Table 1).

The development of exon trapping is based on the work of Cepko *et al.* (24) and others (21, 26, 28), who demonstrated that transmissible retroviral shuttle vectors can be used to generate and recover cDNA versions of defined genomic inserts. Similar strategies have been used to map splice sites in a DNA virus, where spliced products were identified by sequencing or restriction enzyme analysis of all recovered clones (26). This approach is practical only for the analysis of well-characterized genes where transcriptional orientation and exon/intron boundaries are known. In contrast, exon trapping was designed to suggest the presence of genes and recover them from long stretches of genomic DNA. In this scheme, genomic DNA fragments are cloned adjacent to a well-characterized SD and a marked IVS. Potential splicing events are initially identified by a genetic screen in bacteria that detects loss of the marked IVS, and that eliminates the requirement for physical characterization of the insert DNA in every recovered clone. Further analysis is required to demonstrate that a particular recovered clone represents a gene.

Confirmation that a candidate exon is part of a gene and not a consequence of cryptic splicing requires the identification of a transcript. A highly specific exon probe may be recovered from the exon-trap vector by using the PCR. Transcripts can be identified by using this probe to screen Northern blots or cDNA libraries. This probe can also be used to screen a "zoo blot" (1) to determine evolutionary conservation of the putative exon. Such evidence is suggestive of the presence of a gene and is important information if initial screens fail to detect a transcript. In these instances, DNA sequence information obtained during analysis enhances the search for transcripts. For example, a PCR assay based on this sequence can be used to prescreen multiple cDNA libraries or to clone transcripts of very low abundance.

Although we were able to use exon trapping to recover all predicted exons in several different genes, some exons will not be recovered with this method. For example, a small percentage of known genes do not contain introns and therefore will be missed by this screen (23). In addition, some splicing events are temporally regulated or tissue-specific and may not occur in the packaging cell lines used in exon trapping. However, most genes with regulated splicing events also have introns that are constitutively removed, and these would be identified by exon trapping. Finally, because not all DNA sequences are propagated equally well in retrovirus vectors (21), it is possible that the library of recovered clones may not be fully representative of the exons in the starting cloned genomic DNA. Since most genes are composed of multiple exons and the identification of a gene requires the recovery of only a single exon, this consideration should not be a limiting factor.

Exon trapping is a genetic screen that utilizes SA sites as identifiers of candidate exons within cloned mammalian genomic DNA sequences. These candidate exons are ideally suited for establishing the presence of a gene in a cosmid or λ phage insert and facilitating the subsequent isolation of this gene. Our current experience suggests that as many as 20 cosmids can be screened concurrently in a 4-week period. This screen of uncharacterized cosmids will determine the utility of SA signals as identifiers of candidate coding sequences.

We thank R. Cone, R. C. Mulligan, and H. Orr for providing materials and A. Krainer, T. Maniatis, and S. Donner for helpful discussions. This work was supported by grants from the Wills Foundation (to G.M.D. and to R.M.M. and D.R.C.) and from the National Institutes of Health (to R.M.M. and D.R.C.).

1. Monaco, A. P., Neve, R. L., Colletti-Feener, C., Bertelson, C. J., Kurnit, D. M. & Kunkel, L. M. (1986) *Nature (London)* **316**, 336–338.
2. Rommens, J. M., Iannuzzi, M. C., Kerem, B.-S., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J. L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J. R., Tsui, L.-C. & Collins, F. S. (1989) *Science* **245**, 1059–1065.
3. Abe, K., Wei, J.-F., Wei, F.-S., Hsu, Y.-C., Uehara, H., Artz, K. & Bennett, D. (1988) *EMBO J.* **7**, 3441–3449.
4. Spies, T., Blanck, G., Bresnahan, M., Sands, J. & Strominger, J. L. (1989) *Science* **243**, 214–217.
5. Sargent, C. A., Dunham, I. & Campbell, R. D. (1989) *EMBO J.* **8**, 2305–2312.
6. Bird, A. P. (1987) *Trends Genet.* **3**, 342–347.
7. Bell, J. (1989) *Trends Genet.* **5**, 289–290.
8. Gray, M. R., Colot, H. V., Guarente, L. & Rosbash, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6598–6602.
9. Weber, F., de Villiers, J. & Shaffner, W. (1984) *Cell* **36**, 983–992.
10. Allen, N. D., Cran, D. G., Barton, S. C., Hettle, S., Reik, T. & Surani, M. A. (1988) *Nature (London)* **333**, 852–855.
11. Gossler, A., Joyner, A. L., Rossant, J. & Skarnes, W. C. (1989) *Science* **244**, 463–465.
12. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
13. Mann, R., Mulligan, R. C. & Baltimore, D. (1983) *Cell* **33**, 153–159.
14. Miller, A. D., Law, M.-F. & Verma, I. M. (1985) *Mol. Cell. Biol.* **5**, 431–437.
15. Gluzman, Y. (1981) *Cell* **23**, 175–182.
16. Korman, A., Frantz, J. D., Strominger, J. L. & Mulligan, R. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2150–2154.
17. Reed, R. & Maniatis, T. (1986) *Cell* **46**, 681–690.
18. Miesfeld, R., Rusconi, S., Godowski, P. J., Maler, B. A., Okret, S., Wikstrom, A.-C., Gustafsson, J.-A. & Yamamoto, K. R. (1987) *Cell* **46**, 389–399.
19. Koller, B. H. & Orr, H. T. (1985) *J. Immunol.* **134**, 2727–2733.
20. Felgner, P. L., Gadek, T. R., Holm, M., Roman, R., Chan, H. W., Wenz, M., Northrop, J. P., Ringold, G. M. & Danielsen, M. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7413–7417.
21. Brown, A. M. C. & Scott, M. R. D. (1987) *Retroviral Vectors in DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Oxford), Vol. 3, pp. 189–212.
22. Sheffield, V. C., Cox, D. R., Lerman, L. S. & Myers, R. M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 232–236.
23. Smith, C. W. J., Patton, J. G. & Nadal-Ginard, B. (1989) *Annu. Rev. Genet.* **23**, 527–577.
24. Cepko, C. L., Roberts, B. E. & Mulligan, R. C. (1984) *Cell* **37**, 1053–1062.
25. Dougherty, J. & Temin, H. M. (1986) *Mol. Cell. Biol.* **63**, 4387–4395.
26. Dostatni, N., Yaniv, M., Danos, O. & Mulligan, R. C. (1988) *J. Gen. Virol.* **69**, 3093–3100.
27. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 103–119.
28. Sorge, J. & Hughes, S. H. (1982) *J. Mol. Appl. Genet.* **1**, 547–549.