Mini-Exon Epitope Tagging for Analysis of the Protein Coding Potential of Genomic Sequence

BioTechniques 23:116-120 (July 1997)

ABSTRACT

A novel approach to gene discovery and analysis is described. A small exon flanked by consensus 3' and 5' splice sites was synthesized. The exon contains open reading frames encoding 43 amino acid peptides. There are no stop codons in any of the three reading frames, and each reading frame contains an epitope recognized by the same monoclonal antibody. The exon can be inserted into the introns of genes, and the resulting small peptide will be incorporated into the protein encoded by the host gene, regardless of the class of intron. The protein can then be recognized by the antibody, permitting functional studies.

INTRODUCTION

Knowledge of all gene products is an essential component for understanding data obtained from the Human Genome projects. Expressed sequence tags (ESTs) (1) will provide a fragment of transcribed sequence for nearly every gene in the genome, but although they provide suggestive landmarks for positional cloning projects, ESTs usually give little insight into gene function. Furthermore, it is often difficult to forecast the structure of transcripts from genomic sequence, and computational prediction of transcription initiation sites and splice sites continues to be a challenging problem. Enhancer trap experiments (14) give information about the tissue distribution of transcripts but give little data on the function or intracellular localization of protein. Fusion of a marker protein such as β-galactosidase with gene products (4,11) gives more information; however, a disadvantage is that the truncated fusion protein may possess novel properties and may, for example, travel to an inappropriate intracellular compartment.

A method is needed that easily reveals the presence of genes and their functions in a defined genomic interval. This paper presents the development of a mini-exon that can be inserted into introns, permitting the detection of gene products using the same monoclonal antibody regardless of intron class (16). The resultant protein is altered in a minor fashion and can be analyzed using functional assays such as immunofluorescence.

MATERIALS AND METHODS

Construction of the Mini-Exon Oligonucleotide and Related Plasmids

General molecular biology techniques were performed as described (16). The 170-nucleotide (nt) sequence was synthesized using a Cyclone™ Plus DNA Synthesizer (Milligen/Biosearch, Burlington, MA, USA), and because the full-length yield of such a lengthy oligonucleotide is quite low, it was amplified using the polymerase chain reaction (PCR). The reaction contained the 170-mer at 2 mM, and the primers were CTAATTCCTCTCTTCTCCT and CGAGATCTACTTACCTTC, which were at 20 mM each. The conditions were 30 cycles of 94°C for 1 min, 50°C for 2 min and 72°C for 3 min followed by a final extension at 72°C for 15 min.

The PCR products were 5'-phosphorylated using T4 polynucleotide kinase, the 170-bp product purified from a 3% agarose gel and cloned into SmaI-cut Bluescript® KS(+) (Stratagene, La Jolla, CA, USA). A recombinant clone containing the correct insert was identified using double-stranded sequencing and called pmyc2. The mini-exon was excised using KpnI and SacI and ligated into pUC1813 (10), which had been cut using the same enzymes, thus creating pUC1813.myc2. The mini-exon was then excised as a blunt Smal fragment from pUC1813.myc2 and ligated into pBR327 (19), which had been cut with EcoRI and BsaI, the ends having been filled in using Klenow. The resultant plasmid, pBR327.myc2 (2.6 kb), is diagrammed in Figure 1. p11-4cG contains a hybrid cDNA/genomic p53 clone (2,12) borne on a vector containing the
simian virus 40 (SV40) origin, promoter and enhancer. pBR327.myc2 was linearized just upstream of the mini-exon by partial digestion with HindIII and inserted into a HindIII site in intron 9 of the p53 gene in p11-4cG. Since p11-4cG is ampicillin-resistant and pBR327.myc2 is tetracycline-resistant, it was easy to obtain recombinants by selecting for doubly resistant colonies. The resulting plasmid was called p11-4cG.pBR327.myc2.

Analysis of RNA

Total RNA was prepared from COS-1 cells as described (17). Reverse transcription PCR (RT-PCR) was performed as described (18) with minor modifications using the primers recognizing exons 9 and 10 of the p53 gene. The exon 10 primer was also used for RT. For sequencing, RT-PCR products were gel-purified and subcloned into M13mp18. The mini-exon was ligated into M13mp19 for S1 analysis. Preparation of single-stranded radiolabeled probes from the resulting M13 clone and S1 analysis using these probes were as described (17), except the primer used for probe synthesis was the SK primer (Stratagene).

Transfection and Immunofluorescence of COS Cells

The transfection and immunofluorescence of COS cells was performed as described (13) except the NaBH₄ treatment was omitted. Cells were fixed and permeabilized in 4% paraformaldehyde, 1% Triton® X-100, 1× phosphate-buffered saline (PBS) for 10 min, blocked in 1× PBS, 2% bovine serum albumin, 0.1% Triton X-100 for 15 min, incubated in 9E10 antibody (undiluted tissue culture supernatant; Harvard Biolabs Monoclonal Antibody Facility, Cambridge, MA, USA) for 90 min, washed in 1× PBS plus 0.1% Triton X-100, blocked again, incubated for 60 min in fluorescence isothiocyanate (FITC)-conjugated goat anti-mouse IgG (Organon-Teknika/Cappel, Durham, NC, USA) diluted 1:20 in blocking solution, washed and mounted.

RESULTS

The Mini-Exon

The mini-exon (Figure 1) has a 129-bp coding region with no stop codons in any of the three reading frames on the coding strand. In the DNA sequence, there are three direct repeats, which are separated from each other by four nucleotides. Each of the repeats encodes a myc epitope recognized by the same monoclonal antibody (9E10) in one of the three reading frames. Glycine residues flank each of the repeats. These residues are flexible and allow free rotation and mobility of the embedded repeat. Such mobility is thought to enhance antibody binding to epitopes (8). The coding region is preceded and followed, respectively, by a

Figure 1. The mini-exon. (A) The structure of the mini-exon. The branch point sequence (BPS), the polypyrimidine tract [(Py)₁₉], the mandatory AG dinucleotide (underlined) and the cleavage site (3′ CS) of the 3′ splice site are shown. The three ORFs are shown, and the epitope recognized by the monoclonal antibody 9E10 is shown highlighted in each frame. The cleavage site (5′ CS) and the mandatory GT dinucleotide (underlined) of the 5′ splice site is shown. (B) pBR327.myc2. The plasmid is shown in the sense-orientation of the mini-exon, together with useful single cutter sites of the polylinker. The HindIII site was used to introduce the mini-exon into the p53 gene, but it is not a single cutter, since there is also a HindIII site in the tetracycline-resistance (TcR) gene.
consensus 3′ and 5′ splice site (3,15). The 3′ splice site includes a consensus branch point sequence, a polypyrimidine tract and the mandatory AG dinucleotide. The 5′ splice site includes the mandatory GT dinucleotide. Since exon/intron boundaries tend to map to the surface of final protein products (7), the mini-exon peptide will be likely to be displayed on the surface of the protein, where it will be very accessible for binding to antibody (8).

The mini-exon is cloned into a plasmid with the tetracycline-resistance marker, which facilitates further subcloning (see Materials and Methods).

Splicing of the Mini-Exon

To test the mini-exon, it was inserted into intron 9 (class 0) of a mutant allele of the p53 tumor suppressor gene. The recombinant plasmid, p11-4cG.pBR327.myc2, was transfected into COS-1 cells, and RNA was extracted for S1 analysis using DNA complementary to the mini-exon as probe (Figure 2). The observed protected fragment was consistent with correct and 100% efficient splicing of the mini-exon into the p53 transcript. An RT-PCR analysis was performed using primers that lie in the coding region of exons 9 and 10 of the p53 gene and which therefore span intron 9. Figure 3 shows the results of this analysis. If the mini-exon was not included in the p53 transcript, an RT-PCR product of 195 bp was expected; whereas, if it were spliced correctly, a product of 324 bp was predicted. The 324-bp product was observed and was again consistent with correct and 100% efficient splicing. The RT-PCR product was subcloned and sequenced. This revealed that the mini-exon had spliced exactly as predicted into the p53 open reading frame (ORF) (data not shown).

Expression of the Mini-Exon at the Protein Level

COS cells that had been transfected with p11-4cG.pBR327.myc2 were assayed for expression of the hybrid protein by staining the cells with a monoclonal antibody (9E10) that recognizes the repeat present in the mini-exon. This assay revealed that the fusion protein was expressed (Figure 4) and that the protein was present at roughly equal levels in the nucleus and the cytoplasm. The normal location of unmutated p53 is nuclear, although the mutant used here is temperature-sensitive with respect to intracellular localization and is cytoplasmic at 37°C (6). The tagged version is present in the cytoplasm but has regained some of its nuclear localization.

DISCUSSION

Ultimately, computational analysis may permit analysis of the genome with complete confidence, but until then, experiments will be required to deduce gene function. In comparison to other techniques (4,11,14), an advantage of the present approach is that the protein is altered in a relatively minor fashion by the mini-exon and is not truncated. Placing the mini-exon between coding exons is likely to be the least-disruptive position in which to place a new epitope, since exons tend to code for quasi-independent protein domains (7). In addition, the glycines present in the mini-exon will endow the peptide with high flexibility, which may allow the adjacent exons encoded by the gene to assume their natural ori-
entation with respect to each other. The epitope could also be used as a ligand for affinity-purification of the tagged protein using the monoclonal antibody, and this may allow biochemical characterization of the protein’s properties.

Despite the relatively small perturbation of protein structure caused by the mini-exon, it is still possible that the mini-exon could influence a protein’s function or site of intracellular localization, as witnessed by its effect on the mutated p53 gene. In addition, introns do occasionally fall within functional domains. Nevertheless, a broad range of cDNAs have been tagged with the myc epitope, in general, with results faithful to the native protein. The ease of screening for genes with mini-exon epitope tagging (MEET) and the ability to tag multiple introns in the same gene should overcome any disadvantages due to sporadic effects on protein function.

As a tool for gene discovery, the mini-exon could be placed in introns in a random fashion using retroviruses or using bacterial transposons such as γδ, permitting facile detection of proteins. As shown in this work, using the TcR gene, a gratuitous marker can be used, making it simple to detect insertions of the mini-exon in new locations. Since introns make up much of the mammalian genome, there is a large target size for the mini-exon, exceeding that for enhancer trapping. The simplicity of MEET should make it very amenable to automation, which has been one of the driving forces behind recent genomic advances.

Since most mutations affect the protein coding potential of genes, MEET could be particularly useful for mutation hunting by using transfection of tagged clones in conjunction with two-dimensional gel electrophoresis and immunoblotting. In addition, transfection of the mini-exon tagged DNA into cells, combined with immunofluorescence, should give information about the intracellular localization of all gene products in a mutation-critical region. This in itself may provide useful clues; for example, there was a strong belief from the biology of cystic fibrosis that the gene would turn out to be a membrane protein and this presumption proved to be correct (5).

While this manuscript was being prepared, a paper describing a similar approach to gene identification called CD-tagging was published (9). The two methods are conceptually similar, but CD-tagging uses three different vectors or three different antibodies to solve the problem of the three different reading frames by which an intron can interrupt a gene. An advantage of the approach described here is that because only one epitope and cognate antibody can be used universally, a greater speed of screening for new genes should be possible. There is already much experience with the widely used myc epitope/antibody combination, and it displays robust performance capability. Complete sequencing of both the unspliced and spliced exon (see Results) confirms that the epitope should be expressed in all three classes of introns and would be detected by the antibody. Steric hindrance should not be an obstacle to detection of the epitope in the other two frames, since the epitope in the first frame was expressed internally within the p53 gene, but was nevertheless easily detected by the antibody. Furthermore, the use of synthetic consensus splice sequences, as demonstrated in this paper, ensures highly efficient, accurate and constitutive splicing, thus safeguarding against tissue specificity (20).

In conclusion, MEET should be a useful addition to the arsenal of techniques by which the genes uncovered by the genome product can be more accurately catalogued and understood.

REFERENCES


INTRODUCTION

The identification of genes by positional cloning, although laborious, has been facilitated considerably by the construction of framework genetic and physical maps of the human genome. However, the actual identification of candidate gene sequences within a defined portion of the genome remains the most challenging aspect of this process. Consequently, a diverse array

Direct Hybridization of Large-Insert Genomic Clones on High-Density Gridded cDNA Filter Arrays

BioTechniques 23:120-124 (July 1997)

ABSTRACT

A major challenge to positional cloning approaches is the identification of coding sequences within a region of interest. Hybridization of genomic fragments that represent a cloned contig of a defined genomic region on appropriate cDNA libraries theoretically represents a direct solution to this problem. However, this is technically difficult and in general, success with this approach has been limited to the use of small fragments, such as those cloned in cosmids and phages. Since most physical maps are composed of genomic DNA cloned in vectors with significantly greater insert size capacity, there is a need to develop efficient methods to use these clones directly as hybridization probes. Here we describe a highly sensitive protocol for hybridization of P1-derived artificial chromosomes (PACs; average insert size, 120 kb) on a composite, normalized cDNA library comprised of 200 000 clones spotted at high density on nylon filters. Because limited sequence information on more than 150 000 of these clones is now available in the public domain, positive hybridization results can be rapidly converted to cDNA sequence information without recourse to any clone manipulation in the initial phases of a project. Using these protocols, we have been able to reproducibly detect coding exons that constitute as little as 0.2% of the total PAC insert.

INTRODUCTION

The identification of genes by positional cloning, although laborious, has been facilitated considerably by the construction of framework genetic and physical maps of the human genome. However, the actual identification of candidate gene sequences within a defined portion of the genome remains the most challenging aspect of this process. Consequently, a diverse array

Many thanks to Bill Gelbart in whose laboratory this work was performed, to Nick Brown for suggesting the 5′ splice site and to Phil Hinds for p11-4cG. Thanks also to Jan-Fang Cheng, John Morrison and Chris Paszty for helpful comments on the manuscript. I especially thank Jane Iwamura and Sean Egan for encouragement.