Use of PCR to Screen for Promoter Elements in Genomic DNA Library Clones

BioTechniques 26:718-726 (April 1999)

M. Poppe¹, B. Hahm, W. Eickelbaum, M. Arand², N. Paweletz and M. Knehr
German Cancer Research Center, Heidelberg, ¹Interdisciplinary Clinical Research Center, Münster, ²University of Mainz, Mainz, Germany

ABSTRACT

We report a modified PCR strategy to screen for promoter elements of genes of interest that is based upon consecutive rounds of PCR and appropriate subcloning. Following preliminary identification and sequencing of intron 1 by standardized PCR, the application of a suited cDNA/intron primer combination renders a succeeding PCR-mediated screening of cosmids or P1-derived artificial chromosome (PAC) libraries possible, thus identifying genomic clones comprising the searched promoter elements.

We tested our approach in comparison with a commercially available promoter finder kit by searching the promoter elements of the CENP-C gene from the human and mouse genomes. Applying the kit system, we amplified the anticipated promoter from mouse, but failed in isolating human promoter elements. Our approach made use of a 5′-UTR/intron1 primer combination in the second round of PCR, enabling the identification of positive clones from genomic DNA within a human PAC library possible. Subcloning and final PCR amplification revealed the successful isolation of the human promoter.

Therefore, we conclude that our approach might represent a helpful alternative to identify promoter elements, especially when prior art genome walking, STS-based strategies or anchored PCR failed.

INTRODUCTION

The tremendous analytical power of the polymerase chain reaction (PCR) has revolutionized the fields of molecular and cell biology in an astonishing manner and enables the analysis of gene structures and functions with an unprecedented speed and accuracy (11). In particular, the elucidation of promoter elements of genes involved in cell-cycle regulation will enable scientists to get deeper insights into the essential control mechanisms of life and to develop new therapeutic approaches for diseases related thereto. Currently, we are interested in characterizing the regulating sites of the gene encoding centromere protein C (CENP-C), which represents a centromeric protein essentially involved in the correct formation of the kinetochore and which has additionally been shown as involved in the control mechanisms of the important G1-phase of the cell cycle (6). For isolating the CENP-C promoter, we apply current protocols and commercially available kits using the PCR technique. In general, such systems are based upon anchored PCR (4), PCR-mediated genome walking (10, 17) or some sequence tagged site (STS) strategy (9) that makes use of available specific sequence information. However, the latter mostly derives from cDNA sequences, implying one significant disadvantage. Nonfunctional copies of cDNA may exist within the genome as so-called “processed pseudogenes”, which derive from reverse transcription of spliced mRNA molecules into cDNA (19). Since they are nearly homologous to the encoding cDNA of the functional gene, they will be simultaneously amplified by PCR, especially when single copy genes are under investigation (3).

Our approach, as reported herein, overcomes that limitation (Figure 1). Starting by mapping potential splicing sites, which includes testing for consensus sequences of intron/exon boundaries (19), multiple cDNA primer pairs flanking a putative first splicing site are tested in standardized PCR with genomic DNA as a template for the occurrence of an amplification product comprising intron 1 (Figure 1, step 1). Cloning and sequencing of the amplified intron DNA then permits deduction of suitable intron primers, which will be combined with upstream primers deriving from exon 1 or the 5′ untranslated region (UTR) in a further nested round of PCR (Figure 1, step 2). Gene specificity of this second PCR step is a necessary presupposition for the successful screening of genomic libraries for promoter regions. Applying such a pretested gene-specific primer pair in a PCR-mediated genomic screening of, i.e., cosmid or P1 bacteriophage-derived artificial chromosomes (PAC) libraries (5), enables the isolation and characterization of clones potentially comprising the anticipated promoter elements (Figure 1, steps 3–5). Concerning its usefulness and potential limitations, this strategy was tested in comparison with the commercially available PromoterFinder™ DNA Walking Kits to specifically amplify and isolate the promoter elements of the single-copy gene encoding CENP-C from two mammalian species (7,12).

MATERIALS AND METHODS

Enzymes, Chemicals and Kits

Restriction endonucleases, Klenow fragment of DNA Polymerase I, T4 DNA Ligase and the T7 DNA Sequenc-
ing Kit were purchased from Amersham Pharmacia Biotech (Freiburg, Germany); oligodeoxynucleotides and the DIG Labeling and Detection Kit were obtained from Boehringer Mannheim GmbH (Mannheim, Germany); Taq DNA Polymerase was from PE Biosystems (Ueberlingen, Germany); acrylamide, isopropyl β-D-thiogalactopyranoside (IPTG), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and XL1-Blue Competent Cells were from Stratagene (Heidelberg, Germany); agarose was from Life Technologies (Eggenstein, Germany); the QIAEX™ II Gel Extraction Kit was from Qiagen GmbH (Hilden, Germany); Biodyne A Hybridization Filters were from Pall Gelman Laboratory, (Dreieich, Germany); pGEM cloning vector was from Promega (Heidelberg, Germany); pUC18 cloning vector and EcoRI ligation adapters were from Appligene (Heidelberg, Germany); the Human and Mouse PromoterFinder DNA Walking Kits and HeLa genomic DNA were purchased from CLONTECH Laboratories (Heidelberg, Germany); PCR and sequencing primers were synthezised by the German Cancer Research Center DNA Core Facility (Heidelberg, Germany).

**Nested Primer PCR Applying a PromoterFinder Kit and cDNA-derived Primers**

Applying the Human PromoterFinder DNA Walking Kit, a gene-specific primer, HAS1, deriving from the reported cDNA sequence of human CENP-C (12) was tested in combination with a kit adapter primer (AP1) in a first round of PCR (Table 1, B). For that purpose, human genome libraries

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fragment Size</th>
<th>Primer Sequence</th>
<th>Annealing Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) CENP-C, Mouse</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Amplification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adapter Primer 1</td>
<td>No specific band</td>
<td>AP1: 5'-GTAATACGACTCACTATAGGC-3'</td>
<td>60°C</td>
</tr>
<tr>
<td>Gene-specific Primer 1</td>
<td>MAS1: (141) 5'-AGATGGAACATGGCCATGCTTCCGTT-3' (120)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nested Amplification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adapter Primer 2</td>
<td>693 bp</td>
<td>AP2: 5'-ACTATAGGGCCAGCGTGTT-3'</td>
<td>60°C</td>
</tr>
<tr>
<td>Gene-specific Primer 2</td>
<td>MAS2: (119) 5'-CGGCCCAGTTACGTTAAACG-3' (101)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B) CENP-C, Human</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Amplification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adapter Primer 1</td>
<td>No specific band</td>
<td>AP1: 5'-GTAATACGACTCACTATAGGC-3'</td>
<td>60°C</td>
</tr>
<tr>
<td>Gene-specific Primer 1</td>
<td>HAS1: (132) 5'-AATGTCTGGAGTGGAAGCCAC-3' (109)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nested Amplification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adapter Primer 2</td>
<td>1395</td>
<td>AP2: 5'-ACTATAGGGCCAGCGTGTT-3'</td>
<td>60°C</td>
</tr>
<tr>
<td>Gene-specific Primer 2</td>
<td>HAS2: (95) 5'-CAGGAAAGCCAGCAAGACGAAT-3' (72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(C) CENP-C, Human</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Amplification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sense Primer (5'-UTR)</td>
<td>1321 bp</td>
<td>S1: (27) 5'-CTGAGACTTAAGCTTATTGTG-3' (49)</td>
<td>57°C</td>
</tr>
<tr>
<td>Antisense Primer (Exon 2)</td>
<td>AS2: (214) 5'-GTCGACAAAAATCTTCTCTGTAG-3' (192)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second Amplification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sense Primer (5'-UTR)</td>
<td>193 bp</td>
<td>S1: (27) 5'-CTGAGACTTAAGCTTATTGTG-3' (49)</td>
<td>57°C</td>
</tr>
<tr>
<td>Antisense Primer (Intron 1)</td>
<td>AS13: (i34) 5'-CAACCACGCTCGCTTTGACCG-3' (i16)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aAdapter primers derived from PromotorFinder DNA Walking Kits (CLONTECH).

bPosition numbers related to the 5’ end of intron 1. (i).
that had been synthesized by digestion with different blunt-end cutting endonucleases and adapter ligation at the ends of the resulting DNA fragments, served as independent templates in five different reactions (16). PCR was carried out with 1 µL of the different library templates in 20 µL of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 200 µM dNTP, 25 pmol of the appropriate primers and 0.8 U of Taq DNA polymerase, overlayed by 50 µL ofparaffin. The amplification steps were performed with a RoboCycler® Gradient 96 (Stratagene) thermal cycler program as follows: 45 s at 94°C (denaturation), 1 min at 65°C (annealing) and 4 min 30 s at 72°C (extension), performing 35 total cycles of PCR. Following extraction of the amplification products in 70 µL of chloroform/isoamyl alcohol (24:1), 1 µL of each of the products of the five primary PCRs served as a template in a second round of PCR. Applying a nested gene-specific primer (HAS2) in combination with a nested kit adapter primer (AP2), 25 additional cycles of PCR were performed under identical conditions.

Likewise, screening for the promoter elements of murine CENP-C was carried out, applying the Mouse PromoterFinder DNA Walking Kit, together with gene-specific primers MAS1 and MAS2, again in combination with adapter primers AP1 and AP2 (Table 1, A). For this experiment, an annealing temperature of 60°C was chosen in the first round of PCR, which was increased to 65°C in the succeeding nested PCR. Finally, the amplification products of the nested PCRs were analyzed on a 0.7% agarose gel and purified by the QIAEX II Extraction Kit as recommended by the manufacturer.

**Alternative PCR-Mediated Screening Applying a cDNA/Intron Primer Combination**

In a preliminary PCR that was carried out with minor modifications as described before, 1 µg of HeLa genom-
ic DNA served as a template in combination with gene-specific primers S1 and AS2. The latter derived from the 5′ end of the reported cDNA sequence encoding human CENP-C flanking the putative first splicing site (12). Applying thermal cycling conditions of 1 min at 94°C, 1 min at 57°C and 2 min 30 s at 72°C, 45 total cycles of PCR were performed, which were followed by gel electrophoresis and QIAEX II purification of the amplification products as before. Sequence determination of the resulting intron 1 sequence allowed the selection of an intron-derived primer, ASI3, which was combined with cDNA primer S1 in a second PCR under identical conditions (Table 1, C). Since this amplification step appeared to be highly specific, the primer pair S1/ASI3 was chosen for the PCR-mediated screening for the anticipated promoter elements. For this purpose, a PAC library representing the human genome was screened on commission at Genome Systems (St. Louis, MO, USA), again applying identical PCR conditions as we established.

Cloning of PCR Amplification Products and Sequence Analysis

Specific DNA amplification products were extracted from agarose gels and ligated without further processing in the pGEM-T vector followed by transformation of XL1-Blue Competent Cells according to standard methodology (13). Recombinant clones grown as white colonies on LB agar treated with IPTG and X-gal before were amplified as overnight cultures, and recombinant DNA was analyzed following plasmid purification according to the alkaline lysis method (2), and NcoI/NsiI endonuclease was treated to cut out the cloned inserts.

Positive clones deriving from the PCR-mediated screening of the human PAC library were subjected to plasmid purification according to the manufacturer’s protocol (Genome Systems). Subcloning of the PAC inserts was achieved by partial endonuclease digestion with EcoRI, BglII and HindIII, which was followed by treatment of the resulting DNA fragments with Klenow fragment of DNA polymerase I, first ligation to EcoRI adapters, further ligation in EcoRI-cut and dephosphorylated pUC18 vector and transformation in XL1-Blue Competent Cells. Finally, recombinant PAC subclones were analyzed for the presence of possible promoter elements by standard hybridization methodology (13). For this purpose, part of the human CENP-C gene that had been amplified before by PCR, applying gene-specific primers S1 and AS2 (Table 1, C), was labeled by incorporation of digoxigenin and applied as a sensitive probe to screen for subclones comprising the 5′ region. Positive subclones were subjected to DNA sequence analysis, which was performed applying the T7 DNA Sequencing Kit based upon the Sanger method (14). Finally, evaluation of potential binding motifs for transcription factors and partial sequence homologies between the murine and human CENP-C promoter sequences were achieved applying FACTOR and MA-LIGN algorithms within Heidelberg Unix Sequence Analysis Resources (HUSAR) databases.

RESULTS AND DISCUSSION

Identification of the CENP-C Promoter from Mouse Genomic DNA

By applying the Mouse PromoterFinder DNA Walking Kit in combination with cDNA-derived primers MAS1 and MAS2, we obtained a specific 0.7-kb DNA amplification product following the nested PCR step (Table 1, A). Sequence determination revealed 570 bp upstream of the known 5′ end of the reported murine CENP-C cDNA (7), thus possibly reflecting a substantial part of the anticipated mouse promoter. By searching for binding sites of known basal transcription factors, we mapped several regions rich in G/C base pairing (potential SP1 binding sites) and also mapped the putative CCAAT box (Figure 3). Within the murine promoter, we could not map a TATA box, which corresponds to the functions of CENP-C in cell-cycle regulation (6). Essentially, the necessary regulatory elements reflecting the core
Research Reports

promoter site of the CENP-C gene from mouse were successfully isolated using the commercial promoter finder kit, together with cDNA-derived, gene-specific primers.

**Failure in Amplifying the Human CENP-C Promoter by Nested PCR Based Upon cDNA Primers Only**

Likewise, we tried to isolate the human CENP-C promoter sequence applying the Human PromoterFinder DNA Walking Kit in combination with primers HAS1 and HAS2 deriving from the reported human cDNA sequence (12). In spite of the specific amplification of 1.4 kb of DNA (Table 1, B), this amplification product shows the non-functional processed pseudogene of the human CENP-C gene. The latter was confirmed by identifying about 1.3 kb of the encoding sequence of human CENP-C that revealed no intron information at all. In addition, we mapped numerous point mutations, both within the short 5′ UTR and the encoding sequence, one of them representing a premature stop codon at position 78 of the deduced primary sequence.

In principle, amplification of the promoter region deriving from the functional CENP-C gene should have been successful, at least to some minor extent. Within the already known 5′ region of the human cDNA sequence, none of the restriction sites that were used in cloning the promoter finder kits (which might have prevented the amplification of the promoter region) could be mapped. Neither, could specificity of primer annealing be blamed, since the amplified pseudogene sequence of CENP-C represented the only detectable amplification product in gel electrophoresis. Therefore, it seems likely that within the human genome, the number of existing pseudogenes deriving from CENP-C significantly outnumber the two functional alleles. As a result, the dominant pseudogene amplification products revealed by gel electrophoresis gave the impression that the PCR experiment was successful, and consequently, were cloned instead of the anticipated promoter.

**Application of a cDNA/Intron Primer Combination Improves Specificity of PCR Amplification Using Genomic DNA as a Template**

Our approach to identify the human CENP-C promoter by PCR started by establishing some gene-specific intron sequence information to avoid co-amplification or preferred amplification of abundant cDNA copies. Analyzing the 5′ region of the reported CENP-C cDNA (12), we tested cDNA primer combinations in PCR to deduce the first putative splicing site, using primers that were chosen empirically or by taking into account reported consensus sequences describing intron/exon boundaries (19). Applying a combination of primers S1/AS2 with HeLa genomic DNA as a template (Figure 1, step 1), we obtained a specific 1.3-kb amplification product (Figure 2, lane 2). The latter was purified and cloned in pGEM-T/RL1 blue followed by sequence determination, which revealed identity with a minor part of the known cDNA of the human CENP-C gene. However, more important was that we could map the 5′ and 3′ boundaries of the first intron, representing 1133 bp total, which is localized between bases 174 and 175 within the reported cDNA (start codon ATG represents positions 157 to 159). This enabled the design of primers deriving from the 5′ end of intron 1, which were used in a second round of PCR in combination with upstream cDNA primers (Figure 1, step 2). Again in combination with primer S1, use of the antisense intron primer ASI3 (Table 1, C) allowed the specific amplification of the expected genomic 193-bp DNA fragment (Figure 2, lane 3). Thus, we established the presupposition for a successful screening of the searched promoter elements within any genomic library, and in parallel, a probe to immediately test if the CENP-C promoter can be represented in a chosen library or not.

**Identification of the Human CENP-C Promoter by PCR-Mediated Screening of a Human PAC Library**

Amplifying specifically a 193-bp fragment of the 5′ region of the human CENP-C gene, the primer combination S1/ASI3 was chosen to screen a human PAC library under identical PCR conditions (Figure 1, step 3); this was done on commission at Genome Systems. Two PAC clones with insert sizes of

---

![Figure 3](image-url). Comparison of homology of the isolated CENP-C promoter elements deriving from human and mouse genomes. Upstream of the reported cDNA sequences (7,12) (black boxes), about 570 and 620 bp were amplified, either by a commercially available kit (mouse promoter) or by our modified PCR strategy (human promoter), respectively. Partial homologies and localization of CCAAT boxes are emphasized.
120 and 110 kb were positive concerning the CENP-C gene. This was further confirmed by additional PCR testing, again applying the primer combination S1/AS2 (not shown). Following subcloning of the PAC inserts as described (Figure 1, step 4), one of about 500 colonies that were screened with a probe representing intron 1 remained positive in rescreening. Possessing a 3.8-kb insert size (Figure 2, lane 5), the latter was tested in a final round of PCR, repeating once more amplification conditions applying primer combinations S1/AS2 and S1/ASI3 (Figure 1, step 5). Both PCR tests revealed the existence of the 5′ region of the human CENP-C gene, as the resulting DNA pattern confirmed identity with results we obtained earlier (Figure 2, lanes 6 and 7 as compared with lanes 2 and 3). At first, preliminary sequencing from both ends of the insert cDNA did not reveal any known nucleotide sequences from databases. However, the comparison of existing partial homologies to the established mouse promoter elements and the identification of potential binding sites for transcription factors, i.e., regions rich in G/C base pairing and the CCAAT box (Figure 3), supports the successful identification of at least some core region of the human CENP-C promoter.

Advantages and Limitations of the Applied PCR-Based Screening Techniques

Since a preliminary gene mapping had been done, most of the currently existing protocols based on nested PCR or sequence tagged sites (STS) within the genome should guarantee the successful amplification of interesting promoter elements. Comparing the technical side of the PCR strategies we applied, the promoter finder kits and our approach are rather straightforward, convenient and do not require highly sophisticated practical skills. The most important limitation is the necessary specificity of DNA amplification for the gene of interest. Therefore, it is advisable to test different primer combinations beforehand, and then choose the most specific primer pair for the succeeding PCR-mediated screening of genomic libraries.

One major limitation of the PromoterFinder DNA Walking Kits is the maximal length of amplification products one can obtain. Because of the underlying cloning strategy during the production of these kits (8), identification of genomic fragments larger than 3 or 4 kb is rather unlikely. However, since usually essential gene-regulating sites of promoters can be expected within a short range upstream of the translation start site, this should not reflect a serious problem in the majority of experiments. Otherwise, one might prefer the PCR-mediated screening of large cosmid or PAC clones, which thus far have not been digested by restriction endonucleases. Such clones contain much more substantial gene information, which enables the investigation of sequences far more upstream or of other gene sites of interest without requiring a further round of laborious and time-consuming screening.

Our approach, combining consecutive rounds of PCR with well-established DNA purification and subcloning methodology, proved to be also applicable for the successful identification of promoter elements. It is especially applicable if faster protocols using primers only derived from cDNA sequences do not provide successful identification. The preliminary establishing of the sequence of intron 1 is certainly not completely innovative, but it prevents the amplification of cDNA derived from mRNA transcripts (which can happen to some minor extent) or, even worse, the co-amplification of genomic cDNA copies in form of processed pseudogenes. By using the Human PromoterFinder Kit, we indeed amplified the pseudogene sequence instead of the promoter region. Obviously, the application of PCR primers derived exclusively from the cDNA sequence can prevent the successful identification of genomic sequences 5′ of the gene of interest. Besides, the exclusive total use of three gene-specific primers in our approach as compared with the use of only two in other protocols (4,10,17), allows a “real” semi-nested PCR and thus, a “walking-like” approach with improved specificity of priming. A further advantage with our approach is represented by establishing the sequence of intron 1, which permits
the identification of gene-regulating sites therein; this has been reported for an increasing number of genes. It might be difficult to directly achieve the amplification of intron 1. First of all, cDNA base pairs flanking splicing sites frequently do not follow the reported consensus sequences for intron/exon boundaries (19), which makes it advisable to test in parallel with empirically chosen primer combinations to get an idea of the exact localization of intron 1. Secondly, since some genes comprise extremely large intron sizes, as for example the complement factor I gene with 36 kb (18) and the human thrombin receptor gene with 22 kb (15), it might be rather difficult or even impossible to amplify the full length of the first intron sequence. Newly developed expanded PCR techniques (1) applying long PCR kits (i.e., TaKaRa) may help overcome that problem. However, for unknown reasons that might be due to failures of priming in PCR, using that technique, we did not achieve amplification of a genome fragment of CENP-C comprising both introns 1 and 2. Naturally, it is sufficient to determine only a part of the 5′ region of intron 1 to get the information necessary for the design of suitable intron primers.

In summary, we conclude that our approach might be helpful to overcome difficulties occurring with currently available protocols that are not predictable when starting a promoter search. Therefore, we recommend it as an attractive alternative, especially when other PCR-based strategies fail in doing so.

ACKNOWLEDGMENTS

We thank Mr. Werner Fleischer for synthesizing the applied PCR and sequencing primers. Clones Nos. 5386 and 5387 derived from a human PAC library, which was screened on commission by Genome Systems Inc. (St. Louis, MO, USA).

REFERENCES