Degenerate Oligonucleotide Sequence-directed Cross-species PCR Cloning of the BCP 54/ALDH 3 cDNA: Priming from Inverted Repeats and Formation of Tandem Primer Arrays

David L. Cooper and Edward W. Baptist

The Department of Pathology, Duke University Medical Center, Durham, North Carolina 27710

Bovine corneal protein 54 (BCP 54) is the major soluble protein of the bovine cornea, and immunoreactive forms of this protein have been described in a wide range of mammals. Dideoxy sequence determination of a previously synthesized 420-bp cDNA to BCP 54 generated by the novel mixed oligonucleotide primer amplification of cDNA (MOPAC) procedure revealed extensive similarity to the cDNA encoding tumor-associated rat liver (class 3) aldehyde dehydrogenase (RATALD). PCR amplification with additional pairs of degenerate oligonucleotide sequence (DOS) primers derived from both BCP 54-amino-acid sequence and amino acid and nucleotide sequence data from RATALD produced three PCR products that were cloned and subsequently sequenced. The major product was 716-bp BCP 54 cDNA clone encompassing the BCP 54 carboxy-terminal amino acid sequence for which the DOS pair was designed. Sequence alignment of the BCP 54 cDNA and its translation product with RATALD demonstrated 81% and 85% identity at the nucleotide and amino acid levels, respectively. Analysis of the additional two clones established that they were the results of PCR artifactual processes. The first of these was a 552-bp product occurring at elevated primer concentrations that formed through bidirectional amplification from a single DOS annealing to an inverted repeat located in the BCP 54 coding sequence. The second artifactual product was a 212-bp sequence that contained several unreported amplification anomalies, including the formation of a tandem primer array.

Except for the most abundant soluble protein of the bovine cornea (BCP 54, molecular weight 54,000 daltons), little attention has been paid to the water-soluble structural proteins of the cornea. Although the initial description of the major soluble corneal protein can be traced back almost 30 years, BCP 54 had undergone, until recently only minimal molecular characterization with no known function ascribed. We have previously utilized mixed oligonucleotide primers complementary to the reverse translation products of amino acid sequence obtained from Staphylococcus aureus V8-digested BCP 54 fragments and PCR to generate the first reported cDNA probe to BCP 54. The BCP 54 cDNA probe generated by this mixed oligonucleotide primed amplification of cDNA (MOPAC) technique was cloned and dideoxy sequenced. A GenBank library search (version 63.0) revealed a strong similarity to the previously cloned cDNA of rat liver (class 3) tumor-associated aldehyde dehydrogenase (RATALD). Nucleotide and amino acid sequence alignment of the BCP 54 translation product revealed it as 78% and 84% identical with RATALD at the nucleotide and amino acid levels, respectively. Conservation of amino acid elements common to the ALDH supergene family thought to be of structural/functional significance was further substantiated by this analysis.

Degenerate oligonucleotide sequence (DOS) and PCR have been previously utilized successfully to clone cDNAs for specific proteins. This approach relies on the synthesis of oligonucleotides of limited degeneracy from reverse translation of protein subsequences. The cDNAs so generated have been used as probes for Northern and Southern blotting and for screening cDNA and genomic recombinant DNA libraries. DOS primers have also been used to amplify unknown members of gene families and homologous genes in different
been reported concerning the composition, again by the MOPAC technique, of a second extended 716-bp BCP 54 cDNA that encompasses the 3' most terminal protein-encoding region, but additionally discuss the nature of two PCR-generated artifacts associated with the utilization of DOS as PCR primers.

METHODS AND MATERIALS

Materials
PCR primers DOS#1 (AAGCCYTAYGTTGACACARGA) and DOS#2 (GAACTGATRGGCCTTCYTCT) were designed as previously described by reverse translation of amino-terminal amino acid sequences of BCP 54 peptides produced by partial digestion with Staphylococcus aureus V8 protease. Primer DOS#3 (GGTYGGACAARGA) and DOS#2 (GAAC-TGGCATRTRGCG) was previously prepared from reverse translation of the carboxy-terminal 6 amino acids of the RATALD sequence (Fig. 2A). Each degenerate oligonucleotide also contained an 8-nucleotide linker on its 5' terminus that included a 6-nucleotide EcoRI restriction site (GGGAGT). Restriction enzymes and Taq DNA polymerase were purchased from Promega (Madison, WI), and T4 DNA ligase and bacterial alkaline phosphatase from Gibco BRL/Life Technologies (Gaithersburg, MD). Plasmid pBS- was obtained from Stratagene (La Jolla, CA).

PCR Amplification
The first strand cDNA synthesized from bovine corneal epithelium mRNA was used directly in 50-μl PCR reactions which contained 0.2 mM of each dNTP, 200 ng each of primers DOS#1 and either DOS#2 or DOS#3, and 2 units of Taq DNA polymerase in a buffer of 10 mM Tris-HCl (pH 9.0 at 25°C)/50 mM KCl/1.5 mM MgCl2/0.1% gelatin/0.01% Triton X-100. Reactions were overlaid with 50 μl of mineral oil to prevent evaporation and heated at 94°C for 4 min to denature initially the template cDNA. This was followed by 35 cycles of 1 min at 94°C, 1 min at 55°C, and 2 min at 72°C. A final 5-min incubation at 72°C followed to insure completion of all polymerization products. Aliquots (10 μl) of reaction mixtures were analyzed by electrophoresis through 1.2% agarose gels. Following with ethidium bromide and photographed with UV illumination.

A second set of experiments employed identical reaction conditions, except that 400 ng each of primers DOS#1 and DOS#2 or DOS#1 alone were used. This reproducibly generated the 552-bp band that we had previously observed.

Cloning of PCR Products
Reaction mixtures with both the 716-bp and 552-bp bands were extracted with CHCl3 after most of the mineral oil was removed by pipetting. The DNA was precipitated with ethanol, and the pellet was dissolved in 10 μl of 10 mM Tris-HCl (pH 8.5), 1 mM EDTA.

After restriction enzyme digestion with EcoRI, the PCR product was ligated into plasmid pBS- which had been previously digested with EcoRI and treated with bacterial alkaline phosphatase. Transformation of competent Escherichia coli strain NM522 with the ligation reaction mixture resulted in a number of white colonies on ampicillin (100 mg/ml) IPTG (0.5 mM) XGAL (40 mg/ml) LB solid medium. Twelve of these were picked, restreaked, and analyzed by agarose gel electrophoresis of alkaline lysis DNA minipreparations digested with EcoRI.

Sequence Analysis
Plasmid DNA from colonies containing recombinant plasmids was used as template in double-stranded dyeoxy sequencing reactions with T7 DNA polymerase kits from either Pharmacia or US Biochemical. By sequencing from primer sites on both sides of the pBS-polylinker, and the generation of internal deoxylig oligonucleotide primers homologous to the cDNA sequence, the full sequence of the PCR cDNA was obtained. To control for Taq DNA polymerase misincorporations, DNA from 10 independent recombinant isolates identical to clone #6 as described in the text were combined and sequenced as above.

RESULTS
In this report we describe an extended BCP 54 cDNA that encompasses the 3’ most terminal protein encoding region as well as two minor products that were generated during amplification. Our strategy utilizes single-stranded (ss) bovine corneal epithelial cDNA as template for PCR amplification primed by two DOS (Fig. 1) based on previously determined BCP 54 and RATALD amino acid and nucleotide sequence. The 5’ sense primer (DOS#1) and an internal reverse primer (DOS#2) were previously designed by reverse translation of amino acid sequence determined from a BCP 54 V8 protease fragment (Fig. 2A). The 3’ antisense primer (DOS#3) was designed from the 3’ nucleotide and amino acid sequence of the rat homolog RATALD. Based on dinucleotide frequency utilization in human proteins, the 3’ sequence synthesized consisted of the singular methionine codon, the most frequently utilized proline codon, the two most frequent codons of asparagine and alanine, and four of six possible codons of arginine. We also designed a fourth DOS primer to the 5’ terminus of RATALD and paired it with our successful DOS#2 and DOS#3 as well as a number of other internal BCP 54 cDNA-based nondegenerate synthetic oligonucleotide primers. These attempts failed to amplify the 5’ terminus. We believe this may indicate the incomplete nature of 5’ ends contained in our ss cDNA preparation used in this work, and not necessarily an alternative 5’ bovine sequence.

Plasmid DNA from clone #6 which contained a PCR-generated insert of approximately 720 bp (Fig. 2B) was used as template in double-stranded dyeoxy sequencing reactions (Pharmacia or US Biochemical). By sequencing from T3/T7 promoter sites on both sides of the pBS-polylinker, and utilizing two internal synthetic oligonucleotides made to previous BCP 54 cDNA sequence, the full sequence of the PCR cDNA was obtained (Fig. 3). The cDNA contains 716 nucleotides, or, when translated, 239 amino acids. Sequence alignment of the BCP 54 cDNA is colinear with the RATALD sequence without any insertions or deletions. At the nucleotide level, these sequences are 81% identical compared with an 85% identity when translated.
**FIGURE 1** Strategy for the PCR-based amplification cloning of BCP 54 cDNA utilizing degenerate oligonucleotide sequences as primers. First-strand cDNA synthesis was accomplished by AMV reverse transcription of oligo(dT)-primed poly(A)⁺ mRNA isolated from bovine corneal epithelial cells (Fast Track, Invitrogen).

Apparent once more is the striking conservation of glycine residues, as the 5 additional glycines (Gly-403, -404, -409, -411, -415) present in this extended BCP 54 sequence are present in DEHUE1 and RATALD in corresponding positions. The overrepresentation of glycines among residues conserved between distantly related proteins has been correlated with the maintenance of tertiary structure and hints that the tertiary structures formed by the corresponding regions of family members of the ALDH supergene are similar.

Sequencing of clone #3, which contained a 552-bp insert, from the T7 promoter primer revealed complete identity with the 5' end of our previously reported BCP 54 cDNA sequence. However, sequencing from the T3 promoter primer yielded one of the 16 possible sequences of DOS#1 followed by 129 nucleotides of BCP 54 sequence, of which 31 overlap with our previously reported sequence and the remainder of the nucleotides identical to the appropriate region of the clone #6 (716 bp) cDNA sequence reported above. Comparison of the DNA sequence of clone #6 with DOS#1 degenerate sequence of the cloned isolate, since this clone represents only one recombinant isolate, demonstrates that the apparent deoxyoligonucleotide used to prime the amplification of this product, AACCCTCACTATGTCGAAAGA, anneals to the BCP 54 nucleotides corresponding to RATALD nucleotides 1345 to 1357. The 3' terminus of this oligonucleotide exhibits 12 matches and only 1 mismatched pair of bases (Fig. 3), allowing

---

**Sense Primer:** DOS#1

BCP4 amino acid sequence from 33kDa Staphylococcus aureus V8 fragment

\[ \text{GGOACTTACCC(C/T)CA(C/T)TA(C/T)GTG GAC AA(A/G)GA} \]

**Antisense Primer:** DOS#3

Tumor-inducible (class 3) rat aldehyde dehydrogenase amino acid and nucleotide sequence

\[ \text{GGGACTTG(C/T)(C/T)GGG CAT (G/A)TT(G/A)GC} \]

---

**FIGURE 2** Selection of DOS primers, and product analysis in the PCR generation of BCP 54 cDNA. Amino acid sequence is represented by one-letter lowercase code and nucleotide sequence is represented by one-letter uppercase code. (A) The amino sequence (underlined) from the 33-kD S. aureus V8 generated fragment of BCP 54 was used to derive a 5' sense DOS#1 (bold uppercase letter, region of limited nucleotide degeneracy [N/N] indicated) by reverse translation. The antisense DOS#3 was derived from BCP 54 amino acid and RATALD amino acid and nucleotide 3' sequence. The RATALD amino acid and nucleotide numbering are above and below the sequences utilized, respectively. (B) Gel electrophoresis analysis of BCP 54 cDNA PCR products utilizing the DOS#1/DOS#3 and DOS#1/DOS#2 primer pairs. (i) (Lane 1) 1-kb DNA ladder (GIBCO/BRL); (lane 2) production of a single BCP 54 cDNA PCR fragment of 716 bp primed by DOS#1 and DOS#3. (ii) (Lane 1) 1-kb DNA ladder (GIBCO/BRL); (lane 2) optimized production of 420-bp BCP 54 cDNA primed by DOS#1/DOS#2. (iii) (Lane 3) Nonoptimized production of 420-bp BCP cDNA (primed by DOS#1/DOS#2) and 552-bp BCP cDNA (primed by DOS#1).
the BCP 54 sequence in this region to function as an inverted repeat and serve as a second site of DOS#1 binding so that PCR amplification could and did occur with DOS#1 as the sole primer.

We also found that clone #12 contained an unexpected small insert of 212 nucleotides. When this was sequenced, it was obvious that this was a previously unreported type of PCR-generated artifact, consisting of a threefold head-to-tail tandem repeat (tandem array) of primer DOS#2 separated by AC dinucleotides with primer dimer (tail-to-tail) derivatives at each end (Fig. 4).

An earlier report(17) described addition of a single dAMP nucleotide to the 3' end of a blunt-ended DNA duplex. This can be seen at nucleotide 42 of our sequence. Nucleotides 1-41 and loss of the 3'-most nucleotide of primer DOS#2. After the A insertion at nucleotide 94 of our sequence. DOS#1 has undergone a transition to A from the homologous primer DOS#2 in a 2-bp common sequence (nucleotides 103 and 104). At positions 121-122 and 141-142 are insertions of AC dinucleotides between tandem repeats of the DOS#2 primer. The last of these DOS#2 repeats overlaps in its EcoRI linker with the EcoRI linker of a DOS#1 primer. This linker sequence is interrupted by the insertion of a T-A base pair (nucleotide 165) in the TT dinucleotide of the EcoRI site. The final 46 nucleotides are again a classical primer dimer with 2 bp overlap at nucleotides 189-190.

The type of primer dimers represented by nucleotides 43-81 and nucleotides 160-212 has been previously described.(18) We report here two other types of artifactual sequence resulting from primer overlaps. The first of these occurs at nucleotides 26-29 where the 3 pentanucleotide AAAGA of DOS#1 overlaps the 3' pentanucleotide TCTTC of DOS#2. However the 3' C from DOS#2 is deleted and a T is present to pair with the A of DOS#1. It is unlikely that the lack of a C is due to an error in oligonucleotide synthesis because this is the nucleotide that is directly attached to the column matrix. The other type of overlap is at nucleotides 79-81 and at nucleotides 161-166. In both of these, the overlap is between 5' primer sequences which implies that ligation is necessary for unbroken strands.

**DISCUSSION**

This study represents our continued effort to characterize the cDNA encoding the BCP 54 gene product. By use of the PCR and a pair of degenerate oligonucleotide sequence primers derived from both BCP 54 amino acid sequence and sequence data, both amino acid and nucleotide, from the homologous cDNA encoding the tumor-associated rat liver (class 3) aldehyde dehydro-
genase, a second more complete 716-bp cDNA encompassing the BCP 54 carboxy-terminal amino acid sequence was amplified, cloned, and dideoxy sequenced. Nucleotide and amino acid sequence alignment of the BCP 54 translation product revealed it as 81% and 85% identical with RATALD at the nucleotide and amino acid levels, respectively. Conservation of amino acid sequence elements common to the ALDH supergene family thought to be of structural/functional significance were further substantiated by this analysis.

It is apparent that in normal developmental biology, MCP/ALDH 3 is transcribed and expressed in a highly tissue-specific manner, not typical of other ALDH isozymes. Preliminary analysis of the major soluble corneal proteins from other taxa (vertebrate classes: bird, bony fish, reptile, amphibia) found the class 3 ALDH absent in these cornea (22) indicating that BCP 54/RATALD is a taxon-specific protein limited to mammals. However, we did detect enzymatic activity consistent with either class 1 or class 2 ALDH family members (distinct from the class 3 forms) in two species of bony fish. It has also been reported that the taxon-specific crystallin (n-crystallin) of the elephant shrew (19) is a member of the cytoplasmic (class 1) ALDH family, which leads us to hypothesize that the phenomena of gene sharing may be more widespread than originally thought, including the utilization of gene products from alternative members of the same gene family, in different tissues across species boundaries. Consistent with the hypothesis that gene sharing occurs through neutral selection of one of a family of possible genes, corneal expression of BCP 54/RATALD is therefore not based on its enzymatic activity.

We have also analyzed two clones that contained sequences generated partially or completely as PCR artifacts. The larger of these, clone #3 (552 bp), is interesting because it represents an incorrect priming on the desired template. Theoretically, such a PCR product should occur whenever a set of degenerate primers can anneal to a template DNA molecule containing inverted repeat sequences with sufficient sequence similarity to effectively prime Taq DNA polymerase activity. In practice, observation of this artifact was dependent on the relative concentrations of primers and template and not documented until doubling of the DOS primer concentration. Priming from the inverted repeat present within the BCP 54/RATALD protein coding sequence is analogous to single primer PCR, e.g., inter-Alu PCR, (23) a technique used to amplify human DNA from complex mixtures of human and other species DNA.

Our third clone, #12 (212 bp), an obvious product of in vitro processes, is interesting for two reasons. The first is the wide variety of tail-to-tail primer dimers resulting from overlap and subsequent extension of these primers. The second is the formation of head-to-tail tandem direct repeats of one primer separated by dinucleotide inserts that generates a tandem primer array.

ACKNOWLEDGMENTS

This work was supported in part by research grants RO1-EY-00146 and P30-EY-05722 from the National Eye Institute. We thank Mr. Allan Summers for expert technical assistance and Drs. Gordon Klintworth and Joram Piatigorsky for their continuing encouragement of our work in the field of visual science.

REFERENCES

3. Francois, J. and M. Rabaey. 1963. Immunoelectrophoresis of the
proteins of the corneal epithelium. 


Received May 31, 1991; accepted in revised form June 18, 1991.
Degenerate oligonucleotide sequence-directed cross-species PCR cloning of the BCP 54/ALDH 3 cDNA: priming from inverted repeats and formation of tandem primer arrays.

D L Cooper and E W Baptist

*Genome Res.* 1991 1: 57-62

Access the most recent version at doi:10.1101/gr.1.1.57

References

This article cites 20 articles, 3 of which can be accessed free at:

http://genome.cshlp.org/content/1/1/57.full.html#ref-list-1

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here.](#)