PCR-generated artefact from 16S rRNA gene-specific primers

Catherine A. Osborne, Maja Galic, Parveen Sangwan, Peter H. Janssen *

Department of Microbiology and Immunology, University of Melbourne, Parkville, Victoria 3010, Australia

Received 1 March 2005; received in revised form 12 May 2005; accepted 19 May 2005

First published online 9 June 2005

Edited by C. Edwards

Abstract

Artefacts consisting of concatenated oligonucleotide primer sequences were generated during sub-optimally performing polymerase chain reaction amplification of bacterial 16S rRNA genes using a commonly employed primer pair. These artefacts were observed during amplification for terminal restriction fragment length polymorphism analyses of complex microbial communities, and after amplification from DNA from a microbial culture. Similar repetitive motifs were found in gene sequences deposited in GenBank. The artefact can be avoided by using different primers for the amplification reaction.

© 2005 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.

Keywords: PCR artefact; 16S rRNA gene PCR; 16S rRNA gene primers; Primer concatamer

1. Introduction

Knowledge of microbial diversity has increased with the ability to survey genes directly in environmental samples. The 16S rRNA gene has been the target of the majority of such ecological molecular surveys because of its usefulness as a prokaryotic phylogenetic marker [1]. In most cases, the 16S rRNA gene is amplified from total DNA extracted from a sample by use of the polymerase chain reaction (PCR). Slight variants of two primers, 27f and 1492r [2], that allow amplification of nearly complete 16S rRNA genes from the majority of known bacteria, have been used to study bacteria in a diverse range of habitats [1,3]. Rapid community profiling techniques that also rely on PCR amplification, such as terminal restriction fragment length polymorphism (T-RFLP) analysis [4,5] and temperature gradient or denaturing gradient gel electrophoresis [6], are being applied to a wider range of microbial habitats.

Although amplification of bacterial genes in environmental samples has extended our understanding of microbial diversity, PCR has limitations that are still being elucidated. Known problems with PCR include the preferential amplification of certain sequence types, the generation of chimeric sequences, and the occurrence of false positives from experimental contaminants [7–9]. Here we report the generation of a PCR artefact by the commonly used 16S rRNA gene-specific primers, 27f and 1492r, which occurred while trying to amplify 16S rRNA genes from DNA extracted from soil samples, the ilea of mice, and a pure microbial culture. We also report sequences in on-line databases that appear to be the result of this same PCR artefact.

2. Materials and methods

2.1. Generation of T-RFLP profiles

Total community DNA was extracted [10,11] from the ileum of twenty C57BL/6 mice and from eight soil
samples from different sites across Victoria, Australia. We generated T-RFLP profiles targeting the 16S rRNA gene using the oligonucleotide primers FAM27f (5′-GAGTTGTGATCGTCCAG-3′) labelled at the 5′ terminus with 6-carboxyfluorescein, and 1492r (5′-GGTACCTTGTACGAGCTT-3′). T-RFLP analyses were performed as described by Sait et al. [11]. T-RFLP profiles were generated from DNA from the twenty mouse ilea, using HaeIII, and DNA extracted from the eight soil samples, using one of the restriction endonucleases HaeIII, AluI or HhaI (New England Biolabs). Profiles were also generated with DNA extracted from a culture of *Escherichia coli*. Control PCRs were also set up that lacked one critical PCR component (template DNA, dNTPs, Taq DNA polymerase or thermocycling), and then T-RFLP profiles were generated from these. Additional T-RFLP profiles were generated substituting the 1492r primer with either 519r or 1525r [2]. Sterile filter pipette tips were used for setting up all PCRs and T-RFLPs. The water used was double distilled, filter sterilized (0.2 μm, Sartorius Minisart) and UV treated at 254 nm and 1200 mJ cm⁻² for 30 min in a Spectrolinker XL-1000 (Spectronics Corporation). The autoclaved Tris–EDTA buffers [10,11] used to resuspend the sample DNA was filter sterilized and UV treated in the same way. The components of the PCR were UV treated twice at 254 nm and 120 mJ cm⁻² using the “Optimal Crosslink” setting, before addition of the template DNA.

2.2. Generation of sequences

FAM-labelled PCR product generated from one of the soil samples was purified using the Wizard SV PCR Clean Up System (Promega), and approximately 1 ng of the purified labelled product was added as the template in PCRs with unlabelled 27f primer replacing the FAM27f primer. The PCR protocol followed that of Sait et al. [11], except that only two cycles, instead of 25 cycles, were performed. The products from three PCRs were then pooled and purified using the Wizard SV PCR Clean Up System (Promega), ligated into pCR2.1 TOPO vector using the TOPO TA cloning kit (Invitrogen) and transformed into One Shot TOP10 chemically competent *E. coli* (Invitrogen), according to manufacturer’s instructions. White colonies were screened for inserts by PCR using the primers TOP168r (5′-ATGGTTGTGATGAGTTG-3′) and GEM2987f (5′-CCAGTCAC-GACGTTGAAAACG-3′). Nine clones with a variety of insert sizes were sequenced using the T7 primer (Promega) and Big Dye Terminators version 3.1 (Applied Biosystems) and separated using an ABI 3100 DNA sequencer (Applied Biosystems) at the Department of Pathology, University of Melbourne, Victoria, Australia. DNA sequences generated in this study have been deposited in GenBank under the accessionsAY499010 to AY499016.

3. Results and discussion

Profiles generated from DNA from twenty mouse ilea, using HaeIII, and from DNA extracted from the eight soil samples, using any one of three different restriction endonucleases, produced virtually identical T-RFLP patterns (Fig. 1A, B). All of the profiles had an unincorporated primer peak (approximately 24–29 nt in size) and then a peak at 39 nt, followed by peaks corresponding to DNA fragments with further size increments of 41 to 43 nt, regardless of the choice of restriction endonuclease. Undigested PCR ampiclons also yielded the same profiles (Fig. 1C). We attribute this repetitive pattern of peaks to an artefact of PCR (see below). This pattern was detected as the majority of the peaks in some profiles in which non-artefact products were also formed in smaller amounts from mouse ileum community DNA (Fig. 1A) and soil community DNA (not shown).

The artefact was not detected when genomic DNA from *E. coli* was used as the template. Only unincorporated primer was visible in the T-RFLP profiles when the resulting PCR amplicon was not digested, because the expected product was larger than the maximum size detectable using the separation system employed (approximately 500 nt). Peaks corresponding to fragments of 38, 73, and 372 nt were observed after digestion of labelled amplification products from *E. coli* with the restriction endonucleases HaeIII, AluI, and HhaI, respectively. These correspond very closely with the expected fragments sizes of 38, 76, and 372 nt predicted from the *E. coli* 16S rRNA gene sequence (GenBank accession J01695). The amount of product generated in the PCR when using genomic DNA from *E. coli* was in the order of 20 times greater than when mixed community DNA was used as the template, as judged by quantification of PCR products after gel electrophoresis [11]. These results suggest that the artefact is generated when the PCR is functioning sub-optimally. The artefact products may also be generated when the PCR is operating well, but the contribution of these products may have been too small to detect.

When one of two other reverse primers, 519r or 1525r [2], was substituted for the 1492r primer in the PCR with FAM27f, the repetitive peak pattern was not produced from any of the soil and mouse ileum samples (Fig. 1D, E). The artefact pattern was only observed to occur when the 1492r primer was used in conjunction with FAM27f (Fig. 1A–C). Products generated using 519r or 1525r yielded normal non-artefact T-RFLP profiles after restriction digestion (Fig. 1F). The artefact pattern was not observed in samples not subjected to
thermocycling, not containing Taq DNA polymerase, or not containing dNTPs. The repeating pattern was observed in products from two of four PCRs to which no sample DNA was added.

To determine the sequences of the DNA responsible for the 42 nt repeating pattern, FAM-labelled PCR product generated from one of the soil samples was purified as it would be prior to digestion with a restriction endonuclease [11]. Instead of carrying out a digestion, the purified labelled product was added as the template in PCRs with unlabelled 27f primer replacing the FAM27f primer. A clone library was generated from the products of these reactions. Nine cloned fragments of a variety of insert sizes were sequenced. Three of these, approximately 1500 bp in length, had high similarities to legitimate 16S rRNA genes in GenBank databases [12]. The other cloned fragments were 79, 123 (2 clones), 250, 458 and 501 bp long. All of these fragments were made up of repeats of the 27f primer sequence, G, the 1492r primer sequence (reversed and complemented), and AGT (Table 1). The periodicity of the repeating pattern is 42 bp. The sizes of some of these inserts correspond to the sizes of the peaks from the T-RFLP profiles, which must have been produced by concatenation of the primers.

In another experiment [13], we attempted to amplify a 16S rRNA gene from a microbial culture by PCR using unlabelled primers 27f and 1492r. The culture, PV18, was isolated from soil as described elsewhere [13]. The PCR product was cloned and sequenced, as described above, and the sequence was found to be made up of the same repetitive motif found in the T-RFLP products.

Fig. 1. T-RFLP profiles, captured using GeneScan software (Applied Biosystems) and containing the regular repeating pattern of peaks, after digestion with the restriction endonuclease HaeIII of products obtained with the primers FAM27f and 1492r, using: (A) total community DNA from the ileum of a C57BL/6 mouse or (B) total community DNA from a soil sample as the templates for PCR. A profile obtained with undigested PCR products generated using total community DNA from the same soil sample as the template with the primer pair, (C) FAM27f and 1492r also contained the regular peak pattern, but profiles generated with the primer pairs, (D) FAM27f and 1525r, and (E) FAM27f and 519r did not. (F) A normal T-RFLP profile was obtained after digestion with the restriction endonuclease HaeIII of the product obtained with the primers FAM27f and 1525r using total community DNA from the soil sample as the template for PCR. Peaks corresponding to fragments generated by the artefact are labelled with the fragment sizes in panel C.
This indicates that the PCR artefact observed in the T-RFLP profiles does not only occur when one of the primers is fluorescently labelled, but can also occur with unlabelled primers and large amounts of DNA from a microbial culture as the template.

The repetitive sequences generated in this study were found to be very similar to sequences in GenBank databases, from clone library studies and pure cultures (Table 1), showing that this is not an artefact specific to our laboratory. The 16S rRNA gene sequence reported for the pure culture of the Fe-oxidizing bacterium F10 (GenBank accession AY157974; [14]) appears to be made up exclusively of repeats of the 1492r primer sequence, a C, the 27f primer sequence (reversed and complemented), and then AGGT. Several other sequences in the GenBank databases are of partial 16S rRNA genes that apparently have the repeated concatenated primer sequences incorporated into the product. An example (Table 1) is clone A74 (AF477900), generated from a solar saltern [15], which has four full repeats made up of the 27f primer sequence, two other bases (HG), and the 1492r primer sequence (reversed and complemented), before 540 bp that has 99% identity to the 16S rRNA gene of Marinobacter sp. es7 (AJ551128). Some other examples are listed in Table 1. Other occurrences were found in sequences deposited in GenBank. We found no instances in GenBank of a concatemer formed with the primers 27f and 519r, and only one of a concatemer formed with the primers 27f and 1525r.

The generation of PCR artefacts by primer concatamer formation has been shown to affect T-RFLP profiles and amplified 16S rRNA gene sequence data. We are unable to explain why this artefact occurred, but it was found when using normal primers and fluorescently labelled primers, and with DNA extracted from environmental samples and from pure cultures. It did not always occur, even in ostensibly identical experiments. For example, this artefact was not found in experiments using total community DNA from the ilea of a different batch of C57BL/6 mice[11]. The artefact appears to be favoured under conditions where the PCR is functioning sub-optimally, and can be detected by analysing undigested fluorescently labelled products on a sequencing gel. The artefact may not be limited to just one primer set. An artefact like this in a community profiling technique, such as T-RFLP, may lead to erroneous conclusions about the similarity between samples.

Acknowledgements

This work was supported by grants from the Australian Research Council and the Grains Research and Development Corporation.
References