Sensitive and specific cytokeratin 18 reverse transcription-polymerase chain reaction that excludes amplification of processed pseudogenes from contaminating genomic DNA

Peter Tschentscher, Christoph Wagener,* and Michael Neumaier

Processed pseudogenes of residual contaminating genomic DNA interfere with a sensitive detection of cytokeratin 18 (CK18) mRNA by reverse transcription and polymerase chain reaction (RT-PCR). This may cause false-positive results when CK18 mRNA is used as a marker for ectopic tumor cells in specimens from cancer patients. To establish a sensitive CK18 RT-PCR by excluding the amplification of processed pseudogenes, the following strategy was chosen: (a) CK18 pseudogene sequences were cloned from genomic DNA by PCR; (b) cDNA-specific primers were designed on the basis of mismatches between pseudogenes and cDNA; (c) PCR conditions were adjusted to reach maximum sensitivity and specificity. Epithelial cells (1–10) could be detected in 1 mL of blood. Among the numerous CK18 genes homologous to the transcribed gene, at least two different processed pseudogenes exist that are highly homologous to each other and to the exons of the transcribed CK18 gene.

The sensitive detection of tissue-specific mRNA is used increasingly for the detection of micrometastatic tumor cells in peripheral blood, bone marrow, lymph nodes, and other tissues. For some of the target genes, processed pseudogenes are present in the genome. Because the amplification products of cDNA and processed pseudogenes are of the same size, residual contaminating genomic DNA in the RNA preparation may cause false-positive results. Cytokeratin 18 (CK18) is an intermediate-sized keratin-like filament characteristic for epithelial cells. Expression of the human CK18 polypeptide was detected at the protein level in a large number of tumors, cultured carcinoma cell lines, and normal epithelia [1]. To make use of the widespread occurrence of CK18 in carcinoma cells, antibodies against this filament have been applied to the detection of ectopic tumor cells in bone marrow and other specimens [2, 3]. The detection was sufficient in specificity but limited in sensitivity. To lower the detection limit, different groups tried to transfer this diagnostic approach to the nucleotide level detecting CK18 mRNA by polymerase chain reaction after a reverse transcription into cDNA (RT-PCR). They failed for the simple reason that their negative controls, samples without carcinoma cells, turned out to be positive [4, 5]. In contrast to the findings of the protein studies, this outcome was interpreted as nonepithelial expression of CK18, e.g., in hematopoetic cells. In this study, we show that contaminating genomic DNA with processed pseudogene sequences of CK18 interferes with the RNA detection by RT-PCR. To solve this problem, an RNA-specific nested RT-PCR assay was established with the use of sequence information obtained from cloned pseudogene PCR products amplified from genomic DNA.

**Materials and Methods**

PCR

All PCRs were performed in 50 mmol/L KCl, 10 mmol/L Tris-HCl (pH 8.3), 1.5 mmol/L MgCl₂, 0.1 g/L gelatin, 0.25 mmol/L of each dNTP (Pharmacia), 0.4 or 0.8 μmol/L of each primer, and 5 units of Taq polymerase (Life Technologies) in a 100-μL reaction with the use of a Biozym MiniCycler (Hameln) and PCR tubes from Sarstedt. Primers were synthesized with the use of phosphoramidite chemistry on an oligonucleotide synthesizer Gene Assembler Plus (Pharmacia).

Cloning of Pseudogene Sequences

Genomic DNA was prepared from peripheral blood leukocytes of a healthy human volunteer by following established procedures [6] and was treated with 30 ng/L RNase (Life Technologies) for 1 h at 37 °C. For amplifica-
tion of genomic pseudogene sequences, primers were synthesized according to the published sequence of the CK18 cDNA [7]: L, 5′-ATGAGATTCACTGCTGGCTACCTGCT-3′; R, 5′-ATGCTCGAGACTTGGTTGTATCATTG-3′.

RNase-treated DNA (1 μg) served as a template in a PCR with 0.4 μmol/L of primers R and L for 32 cycles at: 97°C (1 min); an annealing temperature of 62°C (2 min), which is optimal according to the Oligo 4.01 program (Primer Analysis Software, National Biosciences, Plymouth, MN); and an extension step at 72°C (2 min), followed by a final extension at 72°C (10 min). The PCR products of five reactions in separate tubes were cloned into the multicloning site of the pCRII vector (Invitrogen). Plasmid preparations of transformed One-Shot-cells (Invitrogen) were checked by restriction fragment analysis to identify clones containing a complete 1.3-kb insert for each of the five initial reactions. The inserts were sequenced with the use of the Tag DyeDeoxy Terminator cycle sequencing system of Applied Biosystems.

RNA EXTRATION AND CDNA SYNTHESIS

RNA was extracted by the acid guanidinium thiocyanate–phenol–chloroform method [8]. To reduce contamination of genomic DNA, an additional step of DNase digestion (Boehringer Mannheim; 40 units in a 400-μL reaction with 10 units of RNase inhibitor in 5 mmol/L MgSO4 at 25°C) was introduced before the RNA was precipitated. The enzyme was inactivated by incubating the RNA preparation for 5 min at 90°C. For measurement of sensitivity of the nested PCR assay, cell culture monolayers of the human colon carcinoma cell line HT29 were trypsinized, pelleted, washed, and resuspended in phosphate-buffered saline, counted in the basophil/lobularity channel of a cell counter (H1, Technicon), and serially diluted (105 to 109/L leukocytes and 153 g/L hemoglobin). After the blood-collection needle was introduced into the volunteer’s vein, the first milliliters of blood were discarded, and the sample was collected in a fresh tube to minimize the risk of contamination with epithelial cells from skin. The dried RNA of 0.5 mL of blood or HT29 cell dilution in blood was resuspended in 80 μL of 10 mmol/L Tris-HCl and 1 mmol/L EDTA (pH 8.0). Random hexamer (Pharmacia)-primed cDNA synthesis was performed in a 20-μL reaction with 10 μL of the RNA preparation and 1 μL of SUPERSCRIPT II RNase H− Reverse Transcriptase (Life Technologies) with the use of buffer and conditions recommended by the manufacturer.

NESTED PRIMERS AND PCR PROTOCOL

With the use of substitutions found by comparison of the highly homologous sequences of the PCR products and the CK18 cDNA, primers were constructed that should preferentially amplify the cDNA: X, 5′-TGTCACACA-CAGTCTGAT-3′; Y, 5′-CAGTCTGAT-3′. Nested PCR was performed with 10 μL of the cDNA with 0.8 μmol/L of primers X and Y in the first PCR of 40 cycles at 95°C (50 s) and 54°C (30 s) and 7 μL of the first PCR product with 0.8 μmol/L of primers X′ and Y′ in the second PCR of different numbers of cycles with the same steps of temperature. Each PCR was started with an initial denaturation of 1 min 20 s at 95°C and terminated with a final extension of 1 min at 72°C. Following established procedures [6] aliquots of the PCR products were electrophoresed in TAE buffer on a 1.3% agarose gel containing ethidium bromide and photographed under ultraviolet light to estimate size and amount.

PATIENTS AND BONE MARROW COLLECTION

Bone marrow samples were collected from five patients undergoing surgery for malignant disease (cancer of esophagus, cardia, and lung) and from one patient with a benign disease (chronic pancreatitis). An incision was made in the skin with a scalpel before the needle was introduced, to minimize the risk of contamination with epithelial cells. Quality of RNA preparation was checked by PCR amplification of β2-microglobulin. Bone marrow samples were obtained by informed consent. The procedures were in accordance with the Helsinki Declaration of 1975, as revised in 1983.

Results

PSEUDGENE INTERFERENCE WITH RNA DETECTION

The analytical problem of processed pseudogene interference with a sensitive detection of CK18 RNA is illustrated in Fig. 1. Conventional primers were designed according to the cDNA sequence without consideration of pseudogene sequences. RNA was prepared from peripheral blood from a healthy human volunteer and transcribed

![Fig. 1](attachment://t.jpg)
into cDNA. In a control experiment with the same RNA preparation the reverse transcriptase was omitted in the cDNA synthesis step. A nested PCR was performed on both samples at optimal annealing temperatures with 40 cycles in the first and 30 cycles in the second PCR step. The transcribed (lane T) as well as the untranscribed sample (lane T') generated a PCR fragment corresponding to the size of 392 bp predicted from the cDNA sequence. The result suggests the following. (a) Residual genomic DNA in the RNA preparation is sufficient to generate a cDNA-independent amplification product. (b) Although the primers span introns, the amplification product is the same size as a cDNA-derived product. Consequently, detection of CK18 RNA cannot be distinguished from amplification of genomic DNA with a conventional nested PCR protocol.

STRATEGY FOR CLONING PROCESSED PSEUDOGENES

The strategy to clone processed pseudogenes by PCR is depicted in Fig. 2: RNA was isolated from a CK18-expressing cell line. DNA was prepared from peripheral blood leukocytes from a healthy human volunteer. After RT of RNA into cDNA, a PCR was performed at an optimal annealing temperature for the primers L and R, which span the whole cDNA as indicated in Fig. 2. The PCR fragment corresponded to the size of 1.3 kb predicted by the cDNA sequence (Fig. 2, lane C). When identical PCR conditions were applied to RNase-treated genomic DNA instead of cDNA, a fragment of identical size was obtained, indicating the amplification of processed pseudogene(s) (lane G). Such products of five independent PCRs were cloned and sequenced. Each of the cloned fragments contained base substitutions when compared with the cDNA. Some substitutions were found in only one clone but not in the remaining clones. These substitutions most probably reflect misincorporations due to Taq polymerase errors. However, a number of identical base substitutions were found at identical positions in more than one clone. Because it is highly improbable that these identical substitutions reflect Taq polymerase errors, they indicate specific differences between the sequences of cDNA and processed pseudogenes. The sequences were aligned and screened for identical base substitutions at identical positions. As shown in Fig. 3, three clones (1, 2, and 5) contained one segment corresponding to clone 3 and a second segment corresponding to clone 4. The finding that different segments of two main sequences are present in a given clone suggests that the five PCR clones may not represent five different pseudogenes. During PCR, an incompletely synthesized DNA strand of one pseudogene may hybridize with another pseudogene in the annealing step of the next cycle. Because each fragment corresponds to either one of the two different clones

![Fig. 2. Strategy for cloning processed pseudogenes.](image)

**Fig. 2.** Strategy for cloning processed pseudogenes.

PCR amplification with primers L and R (annealing temperature 62 °C). Primers are positioned at the 5' and 3' ends of the CK18 cDNA sequence to amplify a full-length cDNA product of 1.3 kb. Lane C, HT29 cDNA; lane N, negative control (no DNA); lane G, RNase-treated genomic DNA.

![Fig. 3. Schematic presentation of CK18 cDNA, intron/exon boundaries, and pseudogene sequences.](image)

**Fig. 3.** Schematic presentation of CK18 cDNA, intron/exon boundaries, and pseudogene sequences.

Arrows indicate the position and size of the introns in the genomic DNA. Light and dark bars of the pseudogenes reflect two sequence motifs with identical base substitutions as compared with the cDNA. Positions of primers X, Y, X', and Y' relative to the cDNA and CK18 gene are indicated.
3 or 4, the five PCR clones may originate from only two different processed pseudogenes.

**DESIGN OF CK18 CDNA-SPECIFIC PRIMERS**

The aligned sequences were screened for regions in which all pseudogene sequences deviated from the CDNA sequence. These conditions were fulfilled for primers designated X, Y, X', and Y'. Additional substitutions were introduced to decrease the stability of primer annealing with the pseudogene sequences. In Fig. 4, the primers X and Y of the first PCR step and the corresponding regions of CDNA and pseudogenes are shown. Primer positions relative to the CDNA and the gene are depicted in Fig. 3. The primers span two introns. From the published sequences, a fragment size of 1054 bp is predicted for the genomic fragment and a size of 413 bp for the CDNA fragment. Because the size of processed pseudogenes corresponds to the size of the CDNA, their amplification would also produce a 413-bp fragment. To prove that the amplification of processed pseudogenes can be prevented with CDNA-specific primers at appropriate annealing temperatures, the RNase-treated genomic DNA of the initial experiment (Fig. 2) was amplified with the use of primers X and Y. As shown in the left part of Fig. 5, fragments of 413 bp are amplified at 63 and 65 °C. The fragment size corresponds to the size predicted from the CDNA. Because the amplification started from genomic DNA, the fragments are derived from processed pseudogenes. With increasing annealing temperature, the only fragment amplified exhibits a size of 1054 bp corresponding to the fragment derived from the CK18 gene. Because the exon sequences of the gene and the CDNA are identical, the experiment indicates that the primers should be suited for a specific amplification of CK18 CDNA even in the presence of processed pseudogenes. To verify this assumption, CK18 CDNA from HT29 cells was amplified at an annealing temperature at which an amplification of processed pseudogenes is excluded (Fig. 5, lanes P1 and P2). Isolated RNA from HT29 cells as well as the RNA prepared from peripheral blood supplemented with these cells produced a fragment of the expected size of 413 bp. Contaminating genomic DNA in the blood sample (lane P2) generated a second fragment of 1054 bp. Because the amplification of processed pseudogenes is excluded at this high annealing temperature, the smaller fragment originates from CDNA and the larger fragment from genomic DNA.

**ESTABLISHMENT AND OPTIMIZATION OF A NESTED PCR PROTOCOL**

The position of primers X, Y, X', and Y' is depicted in Fig. 3. The primers X' and Y' of the second PCR step produce a CDNA fragment of 210 bp and a fragment of 589 bp when the gene is amplified, because the primers span an intron of 379 bp. Taking into account the need of stringent annealing conditions effected by high annealing temperatures, the temperature/time profile of the nested PCR was optimized empirically by detecting RNA of human carcinoma cells HT29, diluted in blood of a healthy human volunteer. A two-step PCR with 50 s denaturation at 95 °C and 30 s annealing at 54 °C without any extra time for extension was most effective. Because of the short annealing time, these conditions favor the generation of short fragments from CDNA vs larger fragments from genomic DNA. To cool the effective temperature in the PCR tube to the desired 70–72 °C during the short annealing step, it was necessary to adjust a lower temperature of 54 °C in the thermal cycler. Omitting reverse transcriptase and RNase treatment of the samples reconfirmed that the produced PCR fragments were strictly RNA/cDNA-dependent and did not originate from processed pseudogenes (not shown).

Applying a constant 40 cycles of this profile in the first PCR meant that the sensitivity of the nested PCR assay was determined by the number of cycles during the second PCR (Fig. 6). As shown on Fig. 6, top, 10² (lane 2) to 10³ (lane 3) epithelial cells per milliliter of peripheral blood could be detected with 25 cycles in the second PCR. With 30 cycles in the second PCR (Fig. 6, middle), 10⁰ (lane 0) to 10¹ (lane 1) tumor cells were detected. The peripheral blood sample without tumor cells remained negative for the CDNA-dependent 210-bp fragment (lane B). The appearance of the genomic fragment of 589 bp...
reflects variable amounts of contaminating genomic DNA in the RNA preparations. Because of a competition between cDNA and genomic DNA, the genomic product, if present in the RNA preparation, becomes apparent only at low amounts of cDNA in the sample (lane 0). However, the genomic amplification product does not disturb a specific detection of the cDNA because the smaller fragment is preferably amplified. When the number of cycles in the second PCR is further increased, a 210-bp product becomes visible even in normal peripheral blood without tumor cells (Fig. 6, bottom, lane B). Corresponding samples with RNase treatment (lane BR) or without reverse transcriptase (lane B) remain negative for this fragment and show the larger fragment of 589 bp, which in this case is not inhibited by competitive amplification of the 210-bp cDNA product. Thus, in contrast to the amplification product of the PCR with conventional primers (Fig. 1), the 210-bp fragment of lane B is strictly dependent on the presence of cDNA, indicating a low concentration of CK18 RNA in the samples (low level transcription).

EXAMINATION OF BONE MARROW ASPIRATES

To prove the diagnostic utility of the CK18 RT-PCR protocol developed in this study, bone marrow aspirates from patients with esophageal, gastric, and lung cancer were analyzed. As negative control, the bone marrow aspirate of a patient with pancreatitis was used (Fig. 7). Seven of 11 samples from patients with malignant disease scored positive, indicating the presence of CK18 mRNA-expressing cells in their bone marrow aspirates. Six results were clearly positive (lanes 5, 6, 7, 9, 11, and 12); in one case only a faint band of expected size could be detected (lane 2). The bone marrow sample from the patient with chronic pancreatitis was negative (lane 1).

Discussion

Previous studies have shown that 15–20 CK18 homologous genes exist in the human genome, including the single active gene and a large number of pseudogenes [9]. Beside those pseudogenes, which retain the exon-intron arrangement of their productive counterpart, a second category has been defined as processed pseudogenes [10] —intronless, nontranscribed genes that are highly homologous to the exons of the transcribed genes. In RT-PCR assays, mRNA is transcribed into cDNA, which is amplified by PCR. Both cDNA and processed pseudogenes lack introns. Furthermore, the DNAs are highly homologous. For this reason, processed pseudogenes may generate PCR fragments of the same size as fragments originating from the cDNA. Thus, processed pseudogenes may present an important analytical problem in RT-PCR assays because residual genomic DNA in the RNA preparation will generate false-positive results.

Cytkeratins have been used widely as protein targets for the detection of micrometastatic cells by immunocytochemical methods [2, 3, 11]. Among the different cytkeratins, CK18 has been the preferred target. In principle, the CK18 mRNA should also be a suitable target for highly sensitive detection of tumor cells by RT-PCR.
However, initial attempts to set up a CK18 RT-PCR failed because positive signals were generated in normal bone marrow and peripheral blood also in the absence of epithelial cells \[4, 5\]. For the following reasons, we assumed that these false-positive results are caused by processed pseudogenes. In Southern transfers, fragments of identical size were generated both from genomic and cDNA \[12\]. As shown here, a PCR product of the expected cDNA size was readily obtained from genomic DNA without RT and after treatment with RNase (Fig. 2, lane G).

The present study was undertaken to establish a CK18 RT-PCR assay for the detection of micrometastatic carcinoma cells. Assuming the existence of processed pseudogenes, primers had to be chosen that would hybridize to the cDNA only. To achieve this goal, it was necessary to provide sequence information on the putative pseudogenes. In principle, this could be done by genomic cloning. To avoid the time-consuming cloning and screening procedures, we took a more direct approach by cloning the genomic PCR products that would disturb a cDNA assay. Experimental procedures were such that only processed pseudogenes were cloned. The sequences of five independent clones were highly homologous, but differed by single base substitutions. Single nucleotide substitutions in one sequence may be caused by misincorporation caused by Taq polymerase errors. However, identical substitutions at the same position in more than one clone most probably indicate a real difference between the amplified genomic sequences. To estimate the number of processed pseudogenes correctly, another possible PCR artifact has to be considered. Incompletely synthesized DNA strands of one pseudogene could hybridize with another pseudogene in the annealing step of the next cycle. For this reason the PCR clones may be composed of parts of different genomic sequences and thus present a false number of processed pseudogenes. However, by focusing on different segments of an alignment, each cloned sequence could be assigned to one of two sequence motifs (Fig. 3). Therefore, the existence of at least two different processed CK18 pseudogenes in the human genome seems reasonable.

On the basis of the substitutions in the pseudogenes, primers were designed that should allow the specific amplification of the CK18 cDNA. To decrease the stability of annealing with the pseudogenes, a few additional mismatches were introduced. At the appropriate annealing temperature, a good amplification of intron-including CK18 gene sequences was obtained from genomic DNA without any amplification of processed pseudogene sequences (Fig. 5, lanes 5–7). Because the CK18 cDNA is identical with the corresponding exons of the active gene, this is a good basis for optimizing a cDNA-specific RT-PCR.

The successful design of cDNA-specific primers on the basis of mismatches to the pseudogene sequences proves the feasibility of the cloning procedure taken in this study. To our knowledge, this is the first report in which processed pseudogenes were cloned by PCR. This direct approach should facilitate the design of cDNA-specific primers for other RT-PCRs in which processed pseudogenes may cause false results.

Time and temperature of the annealing and extension steps of the nested PCR was optimized for the best yield of the desired 210-bp product. A two-step PCR without an extension step and with a short annealing period was most effective because this preferentially amplified small fragments, avoiding competitive synthesis of the intron-including products derived from the active CK18 gene and putative nonprocessed pseudogenes of residual genomic DNA. As a control for RNA specificity, reverse transcriptase was omitted. To reach the required effective annealing temperature within an annealing time of only 30 s, it was necessary to lower the temperature in the thermal cycler to 54 °C.

At a constant number of 40 cycles in the first PCR, the number of cycles in the second (nested) PCR was adopted to reach maximum sensitivity. In reconstitution experiments, 100 carcinoma cells per milliliter of normal peripheral blood (that is, ~17 cells/10⁶ leukocytes) could easily be detected by performing ~25 cycles in the second PCR. This is comparable with the described detection limit of an RT-PCR assay established for cytokeratin 19 \[13\]. Increasing the number of cycles to 30 made possible the detection of 10 to 1 cells/mL of blood (1.7 to 0.17 calculated cells per 10⁶ leukocytes), although in this range specificity became more precarious. With a further increase of cycles, a background RNA signal was detectable also in the sample without carcinoma cells (Fig. 6, bottom, lane B). RNA specificity of the assay was reconfirmed by negative results in the corresponding samples without reverse transcriptase (lane B −) and after RNase treatment (lane B R). These mRNA-based signals may be traced back to a background transcription in normal blood cells. However, although the first milliliter of blood had been discarded and a new collection tube was used when blood was drawn from the cubital vein, contamination of the sample by CK18-positive epithelial cells from skin cannot be formally excluded.

In conclusion, we describe the cloning of CK18-processed pseudogene sequences from human genomic DNA. On the basis of differences in the sequence of pseudogenes and cDNA, a highly sensitive and specific CK18 RT-PCR was developed. In reconstitution experiments, between 1 and 10 epithelial cells could be detected in 1 mL of peripheral blood. In bone marrow aspirates from patients with different carcinomas, positive results were obtained in 7 of 11 samples, and the bone marrow of a patient with nonmalignant disease scored negative. In an ongoing study of bone marrow aspirates from tumor patients, the results obtained by the CK18 RT-PCR compare well with the results obtained by the carcinoembryonic antigen RT-PCR described previously \[14\].
We thank Cornelia Lübcke for her technical assistance in RNA preparation and Jakob Izbicki (Surgical Clinic, University Hospital Eppendorf, Hamburg, Germany) and his coworkers for provision of bone marrow samples from cancer patients.

References