MINIREVIEW

Quantitation of Gene Copy Number and mRNA Using the Polymerase Chain Reaction

(43387)

MATTHIAS VOLKENANDT,*[,1] ADAM P. DICKER,* DEBABRATA BANERJEE,* RENATO FANIN,* BARRY SCHWEITZER,* TETSURO HORIKOSHI,† KATHLEEN DANENBERG,† PETER DANENBERG,† AND JOSEPH R. BERTINO*[,2]

Memorial Sloan-Kettering Cancer Center, New York, New York 10021 and University of Southern California Comprehensive Cancer Center,† Kenneth Norris Jr. Cancer Hospital and Research Institute, Los Angeles, California 90033*

Quantitation of the copy number of a specific gene in a given sample on the genomic as well as on the steady state RNA transcriptional level is of great interest in both basic and clinical research. Traditionally, the copy number of a gene in a given sample is estimated by Southern blot or dot blot hybridization techniques. A certain amount of total genomic DNA of a sample is digested with restriction enzymes, size fractionated by nondenaturing agarose gel electrophoresis, and transferred and permanently bound to a nitrocellulose or nylon membrane (Southern blot); alternatively, DNA can be transferred and bound to a membrane without previous manipulations (dot blot). The membrane is then incubated with a radioactive, labeled molecular probe, which specifically binds to the gene of interest. After washing steps to remove unbound probe, the membrane is exposed to film and a signal is detected. The intensity of the signal, which can be quantitated (e.g., by densitometry), correlates to the amount of specific DNA (gene of interest) per amount of total genomic DNA and gives an estimate of the copy number of this gene.

Similarly, the degree of expression of a gene can be estimated after transferring a certain amount of RNA to a membrane (Northern blot) and hybridization to a specific probe. In each experiment, control samples with a known degree of gene expression have to be analyzed in parallel.

However, there are several limitations of the above techniques.

First, the sensitivity of these assays allows only for detection of gross differences in copy number and expression of a certain gene. Usually, only differences between samples of more than 3- to 4-fold can be detected. Moreover, very low levels of expression of a gene of interest may not be detectable at all.

Second, the amount of DNA or RNA needed for the aforementioned assays is considerable (several micrograms). Quantitative analysis of clinical as well as laboratory samples, from which only minute amounts of nucleic acids can be obtained, is nearly impossible (e.g., analysis of DNA/RNA from small needle biopsies, from formalin-fixed, paraffin-embedded tissue, or, in tissue culture experiments, from single cell colonies).

Third, the techniques mentioned are laborious, and quantitative analyses frequently give ambiguous results. Controls that use DNA or RNA with known levels of amplification or expression of a specific gene must be performed each time for establishment of a standard curve.

Shortly after the advent of the polymerase chain reaction (PCR), the possibility of using this new molecular technique for quantitative purposes was explored by several groups. The PCR is a highly sensitive and efficient method for *in vitro* amplification of a specific gene of interest using genomic DNA as template. Fur-

[1] Present address: Department of Dermatology, University of Munich, Frauenlobstrasse 9-11, D-8000 Munich 2, Germany.

[2] To whom requests for reprints should be addressed at Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021.

P.S.E.B.M. 1992, Vol 200:1-6

0037-9727/92/2001-0001$3.00/0

thermore, after synthesis of a cDNA using reverse transcriptase, specific RNA molecules may also be amplified *in vitro*. Due to a cyclic change in temperature, an exponential amplification of DNA or cDNA molecules of interest occurs, doubling the amount of these molecules during each cycle (1, 2). After PCR amplification, the reaction product can be electrophoresed on an agarose gel and a signal representing the amplified gene of interest can be visually detected by ethidium bromide staining. Minute amounts of DNA or RNA are sufficient to yield a clearly detectable signal. Therefore, the potential of this method for quantitative analysis of a gene or gene expression from small numbers of cells or small tissue samples is enormous.

However, as a potential limitation of quantitative PCR analyses, it was noticed that repetitive experiments, using both identical PCR conditions as well as identical amounts of the same template DNA, did not necessarily lead to identical amounts of amplification products. For example, when a master mix containing all the reagents as well as template DNA for PCR was prepared and then split into several separate tubes, the yield of the amplified product varied as much as 6-fold (3). An explanation may be that the efficiency of PCR is affected even by minute and uncontrollable differences in concentrations of reagents present in the sample (DNA, dNTP, primers, $MgCl_2$, Taq polymerase), as well as by small differences in temperature depending on the position of the reaction tube in the heat block of the thermal cycler (4). Thus, by itself, PCR did not seem to be a reliable quantitative assay.

Quantitative PCR by Co-Amplification of an External Standard

At one of the first meetings on PCR technology, Gilliland *et al.* (5) presented an approach that potentially obviates these problems. The strategy involves co-amplification of a competitive synthetic template ("external standard"), which is added to the reaction mixture and is amplified in the same tube and with the same pair of primers as used for amplification of the target gene. As differences in reaction conditions will equally affect amplification of competitive template as well as target gene, the relative ratio of both amplification products will remain independent of variations in conditions during amplification. Competitive template DNA is added in various and known amounts to several PCR mixtures with the same amount of target gene. Comparing the intensity of the two amplified products, an initial concentration of competitive template is determined that yields an identical amount of amplification product as the target gene (ratio of competitive gene to target gene, 1). In this tube, starting concentrations of competitive and template DNA are identical and the initial concentration of the target gene can be determined.

As an example, the genomic copy number of the granulocyte-macrophage colony-stimulating factor (GM-CSF) gene was quantitated in a number of cell lines. Primers used for PCR amplification annealed to two different exons flanking a small intron. For a competitive template, a vector containing GM-CSF cDNA was chosen. Thus, amplification occurs by using the same pair of primers, but products are distinguishable by size (with and without an intron). By titrating equal amounts of genomic DNA (1 μg) against a series of dilutions with known amounts of plasmid GM-CSF cDNA, the genomic gene copy number of GM-CSF may be determined exactly. In their experience, even patients with $5q^-$ syndrome (having only one copy of the GM-CSF gene) could be distinguished from normal individuals (with two copies of this gene) (3).

Similarly, expression of GM-CSF could be analyzed by quantitation of GM-CSF mRNA molecules. GM-CSF cDNA synthesized from RNA of equal numbers of cells was co-amplified with a dilution series of known concentrations of genomic plasmid GM-CSF DNA. Cells stimulated for GM-CSF expression could be distinguished from unstimulated cells and the number of GM-CSF mRNA molecules could be estimated. Alternatively to competitive molecules that differ by size, a mutant competitive template, which differs from the target cDNA sequence by a new restriction site, can be used (6). After amplification, products of competitive and target template are distinguished by restriction enzyme digestion. This has the theoretical advantage that conditions for amplification of target and competitive DNA are almost identical since they do not differ in size and overall sequence (3, 6).

Using a similar approach, Wang *et al.* (7) quantitated specific mRNA molecules of various cytokines. They designed a vector containing sequences complementary to specific primers used for amplification of various cytokine cDNA. This "multiple primer region" is flanked upstream by the T_7 polymerase promoter and downstream by polyadenylated sequences. The cDNA can then be converted to cRNA with T_7 RNA polymerase. A known amount of purified cRNA molecules are added to a certain amount of RNA of a sample to be analyzed; both are co-reverse transcribed into cDNA and serial dilutions of the reverse transcriptase product are amplified for a specific cytokine. The variable template concentrations of the internal standard cRNA and the sample RNA are plotted against the intensity of their amplification products and the amount of target mRNA can be extrapolated by comparison of the two curves.

Our interest focused on the analysis of the copy number for the human dihydrofolate reductase (DHFR) gene. Methotrexate, a folate analog, binds to DHFR and inhibits its activity, thus leading to a depletion of the reduced folate pool in cells, which in turn

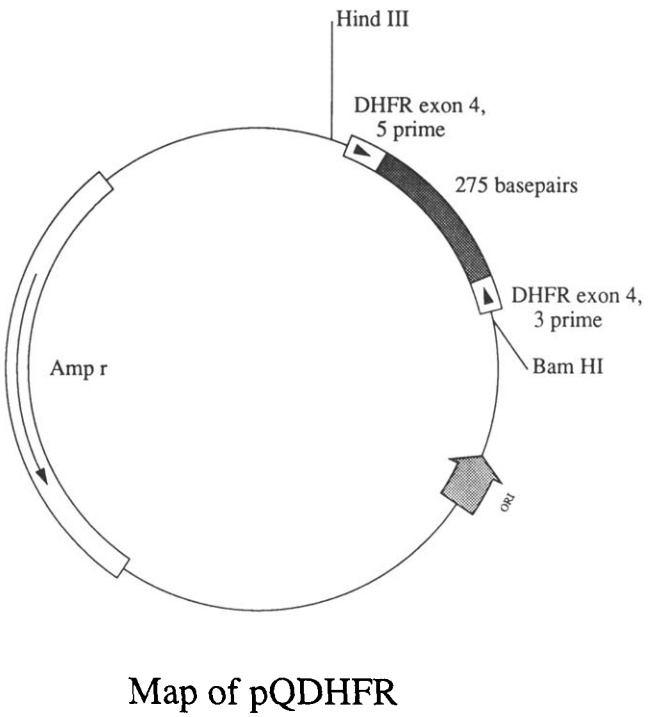
leads to inhibition of thymidylate and purine nucleotide biosynthesis. An increased level of DHFR gene copy number, resulting in an increased level of the enzyme, is a mechanism of acquired methotrexate resistance of tumor cells. This has been shown frequently in tissue culture experiments, but only in very few cases in patients (8). However, a very low level of DHFR amplification (2- to 3-fold) may not be detectable by conventional methods, but may play a role in the clinic.

We designed two primers for PCR amplification of exon 4 of the DHFR gene. These primers amplify a 183-base pair segment. For co-amplification, we constructed an artificial DNA segment, which was ligated into a plasmid vector (pQDHFR). This insert consisted of 275 random nucleotides, flanked on each side by 20 nucleotides complementary to sequences of the two primers used for amplification of exon 4 (Fig. 1). As a result, this insert will be co-amplified by these primers, yielding a 315-base pair product (275 + 20 + 20 = 315 base pairs), distinguishable from exon 4 by gel electrophoresis.

For quantitation of DHFR gene copy number in a given sample, a PCR mixture of 630 μl containing 7 μg of sample DNA was prepared and aliquoted into seven tubes (each 90 μl will contain the same amount of sample DNA, 1 μg). We then added 10 μl of various dilutions of competitive vector DNA, and PCR was performed.

Figure 2 shows the results of amplification. The upper bands represent the amplification products of competitive vector DNA, which was added in decreasing concentrations to the sample before PCR. The lower bands represent the amplification products of DHFR exon 4. Although each test tube contained the identical amount of genomic DNA (1 μg), there was an increasing intensity of the amplification product, due to decreasing competition by vector DNA in the amplification process. The concentration of each amplification product can be determined by densitometric scanning of negatives of photos taken from ethidium bromide-stained gels, or by excision and scintillation counting of bands that are radioactive because of incorporation of [α-^{32}P] dATP. The ratio of intensities of amplified products can be determined (competitive gene to target gene), and at a ratio of 1, the number of molecules in the initial sample is equal for both genes.

However, in our experience, this assay is not without problems. While the competitive DNA usually was amplified with an efficiency correlating to the amount of vector DNA in the initial sample, amplification of genomic DNA frequently did not yield products of gradual increasing intensity (Fig. 3). Thus, although the external standard is amplified with the same pair of



Map of pQDHFR

Figure 1. Diagram of vector pQDHFR used as competitive template for PCR amplification reactions of exon 4 of the human dihydrofolate reductase gene. When 183 base pairs of exon 4 of the DHFR gene are amplified, 315 base pairs of the vector (275 + 40 base pairs representing primer sequences) are co-amplified in the same tube using the same pair of primers.

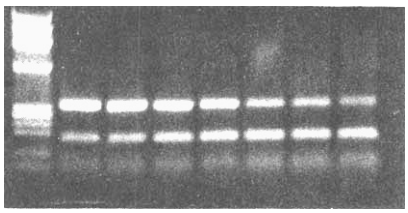


Figure 2. Quantitative analysis of the human DHFR gene using external vector DNA as template for co-amplification. The upper bands represent the amplification products of competitive vector DNA, which was added in decreasing concentrations to the sample before PCR. The lower bands represent the amplification products of DHFR exon 4. Although each tube contains the identical amount of genomic DNA (1 μg), there is an increasing intensity of the amplification product of DHFR.

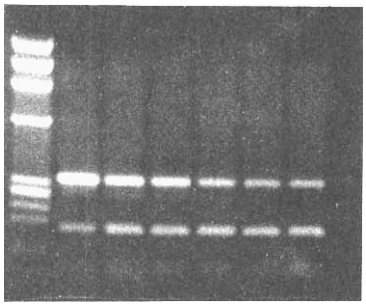


Figure 3. Quantitative analysis of the human DHFR gene using the same conditions as in Figure 2. A decrease in amplification product of competitive vector DNA is seen; however, no continuous increase in amplification product of DHFR occurs.

primers under presumably identical conditions, this technique is associated with possible inherent limitations.

1. The external standard, in our experience, seems to have an "amplification advantage" over the target gene and frequently is amplified more efficiently. This is probably due to a much less complicated tertiary structure of the vector as compared with genomic DNA. This may be less of a problem when amplification of target cDNA and vector DNA is compared, or when the genomic DNA is digested with a restriction enzyme prior to PCR amplification.

2. When analyzing RNA from clinical samples, many circumstances may affect the RNA of the sample, but not the competitive template (RNA degradation, possible effects of anticoagulation used, etc.).

3. As several reactions (dilution series) are required for quantitative analysis of each individual sample, a relatively large amount of DNA is needed (at least 5 μg). Quantitation of gene copy number of DNA from small clinical samples or from DNA from tissue embedded in paraffin will not be possible.

4. The concentration of sample DNA as well as competitive template before PCR has to be known precisely, and small sampling or measurement errors will drastically affect the validity of the results.

5. A vector has to be constructed for each gene to be analyzed. Once this is accomplished, the assay is still very labor and cost intensive.

Quantitative PCR by Co-Amplification of an Internal Standard

As an alternative approach to quantitative analysis of a target gene by PCR, a *single copy* endogenous reference gene (internal standard) can be co-amplified in the same reaction tube by using an additional pair of primers. The copy number of the target gene in the initial sample is estimated by comparing the intensity of the amplification products of the target gene and the reference gene. This method, called differential PCR, was described by Frye *et al.* (9) and used for quantitative evaluation of *neu* proto-oncogene amplification in breast carcinoma cell lines. As the single copy reference gene, part of the interferon-τ gene was co-amplified in the same tube. When placental DNA or DNA from cell lines not amplified for the *neu* oncogene was analyzed, the ratio of the concentration of amplification products (*neu* to interferon-τ) was between 0.6 and 1.5. The ratio in all cell lines with known amplification of the *neu* oncogene was clearly higher. Even cell lines with known 2-fold amplifications could be identified. However, a linear correlation between the *neu* to interferon-τ ratio and the level of amplification of the *neu* gene could not be established.

Noonan *et al.* (10) used a similar technique to quantitate the expression of the multidrug resistance gene (MDR1). After reverse transcription of RNA from clinical samples, specific MDR1 cDNA sequences as well as cDNA sequences of the endogenous β2-microglobulin gene (β2-m) were amplified *in vitro*. Over a certain number of PCR cycles (which depend on the level of MDR expression and have to be determined for each individual sample), the ratio of intensities of MDR1 to β2-m-specific PCR products showed a linear increase when plotted against MDR1 copy number per cell.

For estimation of the copy number of the DHFR gene, we amplified 220 base pairs of exon 3. As internal standard, 120 base pairs of intron sequences of the β2-m gene were co-amplified. Exon 3 was found to be amplified with slightly higher efficiency than exon 4, when β2-m was co-amplified. However, using placental DNA as template, primers for DHFR still amplified much less efficiently than those used for β2-m, when equimolar concentrations of primers were used. Consequently, we determined concentrations of primers that yielded amplification products of equal intensity (ratio of DHFR to β2-m $\approx$ 1). This was achieved when, on a molar basis, approximately four times less primers for amplification of β2-m were used. When DNA from cell lines with increasing levels of DHFR copy number was used as the template for amplification, the ratio of DHFR to β2-m increased, both due to an increase in DHFR as well as to a decrease in β2-m amplification product (Fig. 4). In another experiment, a different gene was used as an internal standard for co-amplification (120 base pairs of the N-*ras* oncogene) and 183 base pairs of exon 4 of DHFR were amplified. When using DNA from six cell lines with increasing levels of DHFR copy number (1, 2.8, 4.6, 6.4, 8.2, and 10), an increase of the ratio of DHFR to *ras* was observed (Fig. 5).

A disadvantage of using an internal standard for quantitative PCR is the need for a second pair of primers for co-amplification of the endogenous internal standard gene. Under conditions used for amplification of the target gene, another pair of primers may amplify

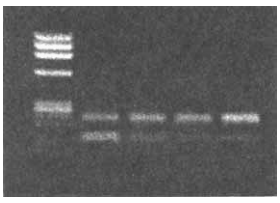


Figure 4. Quantitative analysis of human DHFR gene using sequences of an endogenous gene as template for co-amplification (internal standard). The upper bands represent amplification products of DHFR exon 3 (220 base pairs). As internal standard, 120 base pairs of the β2-microglobulin gene were co-amplified (lower bands). Analysis of four cell lines with increasing levels of DHFR copy number (from left to right) yielded an increase in the ratio of amplification products (DHFR to β2-m).

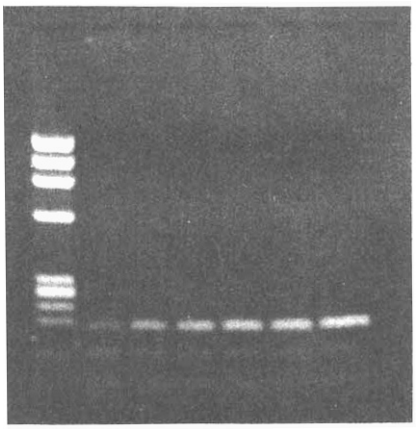


Figure 5. Quantitative analysis of the human DHFR gene using 120 base pairs of the N-*ras* oncogene as template for co-amplification. The upper bands represent amplification products of DHFR exon 4 (183 base pairs). Analysis of six cell lines with increasing levels of DHFR copy number yielded a linear increase in the ratio of amplification products (DHFR to N-*ras*).

with different efficiency. However, primer concentrations can be determined that lead to amplification products of equal intensity.

In comparison to methods using an external standard, quantitative PCR by co-amplification of an internal standard is characterized by the following features: (i) Only one reaction is required for quantitative analysis of an individual sample; (ii) Small amounts of template DNA are sufficient for analysis; (iii) Small variations in concentrations of template DNA will not be of crucial importance, as target DNA as well as the co-amplified internal standard gene will be equally affected; (iv) No vector has to be constructed as a competitive template. If for some reason amplification of a specific sequence of the target gene does not work efficiently, other sequences can be amplified by designing another pair of primers (as opposed to construction of a new vector); and (v) While low levels of increase in copy number of the target gene can be identified, our experience suggests that the level of amplification can only be estimated and not be determined exactly.

A novel approach (11) uses the advantages of the strategy in which an internal endogenous standard gene is amplified, while the problems associated with two amplification reactions in one tube using two different pairs of primers are avoided.

To the 5′ end of the 5′ primer, which is used for PCR, sequences of the T_7 promoter are added. After PCR amplification, an aliquot of the product is then transcribed into RNA in the presence of [α-^{32}P]dATP and the transcription product is electrophoresed on a denaturing polyacrylamide gel. The radiolabeled RNA is then excised from the gel and quantitated by scintillation counting. In preliminary experiments, amplification reactions using amounts of template DNA differing over a wide range (several logs) are performed. For all genes tested so far, a small range of concentrations of template DNA (differing by approximately one log) was found to consistently lead to a linear increase of intensities of the amplification products. This range is constant for a specific gene and for specific amplification reactions (including the pair of primers) used. The concentration of template DNA yielding a linear increase in the amount of amplification products is

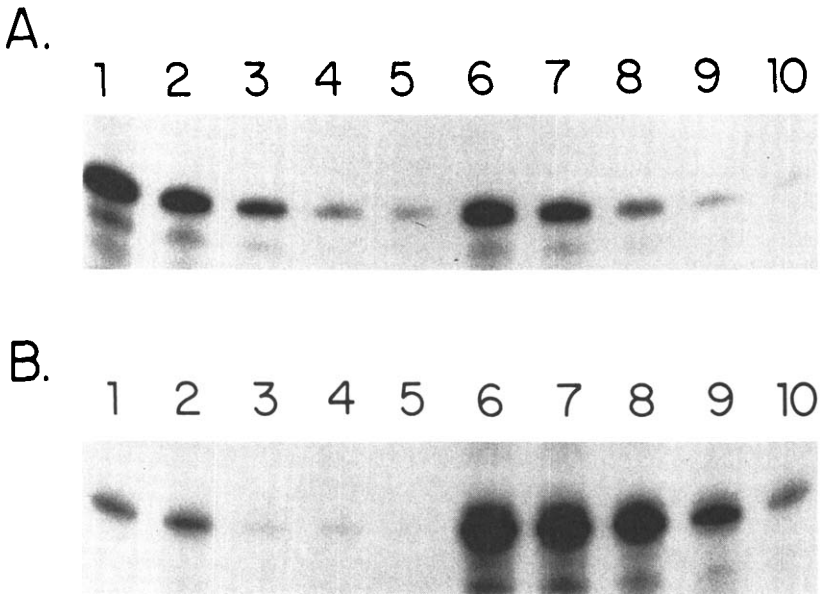


Figure 6. Quantitative analysis of the human cell line CEM/S and cell line CEM/E with a known 10- to 15-fold amplification of the DHFR gene. Various amounts of DNA were PCR amplified for β-actin (A) and for DHFR (B) and transcribed into RNA. CEM/S, reactions 1–5; CEM/R, reactions 6–10.

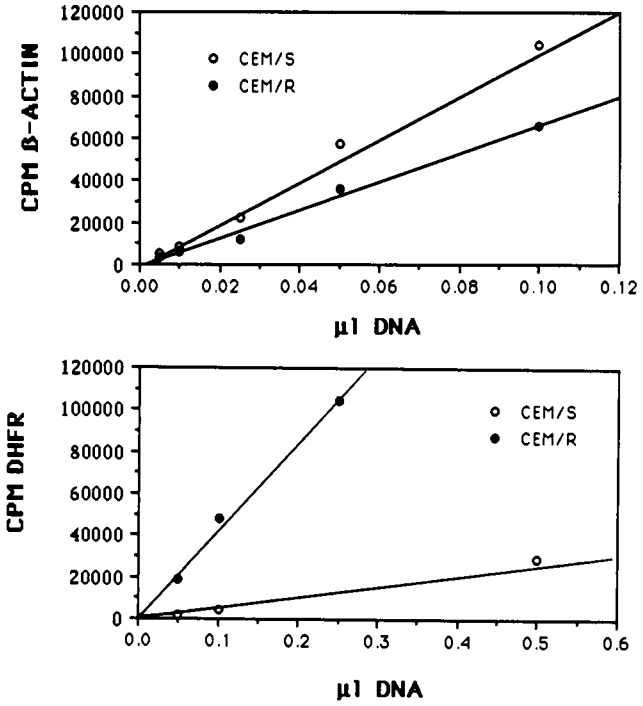


Figure 7. Radiolabeled RNA (Figure 6) was excised from the gel and quantitated by scintillation counting. This figure shows graphical representation of the data (cpm/μl DNA solution). Using these linear curves, the intensity of RNA transcripts of PCR products generated by 1 μl of DNA solution was determined by extrapolation. β-actin (CEM/S), 1,077,326; β-actin (CEM/R), 672,082; DHFR (CEM/S), 60,525; DHFR (CEM/R), 417,696. The ratio of DHFR to β-actin was determined (for CEM/S, 0.056; for CEM/E, 0.62) and revealed an 11-fold difference in DHFR gene copy number in these cell lines.

usually exceedingly low (between 0.01 ng and 40 ng) and visualization and quantitation of the products are possible due to transcription into RNA (leading to a 500- to 1000-fold further increase in yield).

With this method, the target gene and the endogenous standard gene are amplified separately in individual tubes (11). This has the advantage of allowing for the determination and the use of different optimal amplification conditions for each gene (temperatures, amount of template). This is an advantage, since optimal amplification conditions (yielding a linear increase of amplification products) may differ between genes and primers used for amplification.

When analyzing gene copy numbers in an unknown sample, sequences of the target gene as well as the endogenous standard gene (e.g., β-actin) are amplified in separate tubes using about four different amounts of DNA as template that have been shown previously to lead to a linear increase in intensities of amplification products. Using these linear curves, intensities of PCR products generated by the same volume of DNA solution are determined by extrapolation (e.g., for 1 μl of template DNA solution) and the product ratio (target gene to internal standard gene) is determined. Using this strategy, two cell lines with a known 10- to 15-fold difference in DHFR gene copy number were analyzed, and the difference was clearly determined (Figs. 6 and 7).

The same approach can also be used for quantitation of the relative amounts of gene expression, if the total RNA isolated from cells or tissues is converted to cDNA.

These developments promise to make quantitative PCR analysis an exciting new technique applicable to a wide array of questions both in clinical as well as basic research.

Matthias Volkenandt is supported by Grant Vo 415/1-1 of the Deutsche Forschungsgemeinschaft. Joseph R. Bertino is an American Cancer Society Professor. This study was supported by NIH Grant CA08010 and by ACS Grants BC 561-C and CH-1.

1. Eisenstein BI. The polymerase chain reaction. A new method of using molecular genetics for medical diagnosis. N Engl J Med **322**:178-183, 1990.
2. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487-491, 1988.
3. Gilliland G, Perrin S, Blanchard K, Bunn HF. Analysis of cytokine mRNA and DNA: Detection and quantitation by competitive polymerase chain reaction. Proc Natl Acad Sci USA **87**:2725-2729, 1990.
4. Linz U. Thermocycler temperature variation invalidates PCR results. Bio Techniques **9**:286-293, 1990.
5. Gilliland G, Perrin S, Bunn F. Quantitative amplification of mRNA using polymerase chain reaction [Abstract]. J Cell Biochem [Suppl] **13E**:270, 1989.
6. Becker-Andre M, Hahlbrock K. Absolute mRNA quantification using the polymerase chain reaction (PCR). An novel approach by a PCR aided transcript titration assay. Nucleic Acids Res **17**:9437-9446, 1989.
7. Wang AM, Doyle MV, Mark DF. Quantitation of mRNA by the polymerase chain reaction (published erratum appears in Proc Natl Acad Sci USA 87(7):2865, 1990). Proc Natl Acad Sci USA, **86**:9717-9721, 1989.
8. Schweitzer BI, Dicker AP, Bertino, JR. Dihydrofolate reductase as a therapeutic target. FASEB J **4**:2441-2452, 1990.
9. Frye RA, Benz CC, Liu E. Detection of amplified oncogenes by differential polymerase chain reaction. Oncogene **4**:1153-1157, 1989.
10. Noonan KE, Beck C, Holzmayer TA, Chin JE, Wunder JS, Andrulis IL, Gazdar AF, Willman CL, Griffith B, Von Hoff DD, Roninson IB. Quantitative analysis of MDR1 (multidrug resistance) gene expression in human tumors by polymerase chain reaction. Proc Natl Acad Sci USA **87**:7160-7164, 1990.
11. Horikoshi T, Danenberg KD, Stadlbauer THW, Volkenandt M, Shea LCC, Aigner K, Gustavsson B, Leichman L, Frosing R, Ray M, Gibson NW, Spears CP, Danenberg PV. Quantitation of thymidylate synthase, dihydrofolate reductase, and DT-diaphorase gene expression in human tumors using the polymerase chain reaction. Cancer Res (in press).