A new mathematical model for relative quantification in real-time RT–PCR

Michael W. Pfaffl*

Institute of Physiology, FML-Weihenstephan, Center of Life and Food Sciences, Technical University of Munich, Germany

Received December 18, 2000; Revised February 21, 2001; Accepted March 14, 2001

ABSTRACT

Use of the real-time polymerase chain reaction (PCR) to amplify cDNA products reverse transcribed from mRNA is on the way to becoming a routine tool in molecular biology to study low abundance gene expression. Real-time PCR is easy to perform, provides the necessary accuracy and produces reliable as well as rapid quantification results. But accurate quantification of nucleic acids requires a reproducible methodology and an adequate mathematical model for data analysis. This study enters into the particular topics of the relative quantification in real-time RT–PCR of a target gene transcript in comparison to a reference gene transcript. Therefore, a new mathematical model is presented. The relative expression ratio is calculated only from the real-time PCR efficiencies and the crossing point deviation of an unknown sample versus a control. This model needs no calibration curve. Control levels were included in the model to standardise each reaction run with respect to RNA integrity, sample loading and inter-PCR variations. High accuracy and reproducibility (<2.5% variation) were reached in LightCycler PCR using the established mathematical model.

INTRODUCTION

Reverse transcription (RT) followed by the polymerase chain reaction (PCR) is the technique of choice to analyse mRNA expression derived from various sources. Real-time RT–PCR is highly sensitive and allows quantification of rare transcripts and small changes in gene expression. As well as this, it is easy to perform, provides the necessary accuracy and produces reliable as well as rapid quantification results. The simplest detection technique for newly synthesised PCR products in real-time PCR uses SYBR Green I fluorescence dye that binds specifically to the minor groove double-stranded DNA (1). The quantification method of choice depends on the target sequence, the expected range of mRNA amount present in the tissue, the degree of accuracy required and whether quantification needs to be relative or absolute (2). Generally two quantification types in real-time RT-PCR are possible. (i) A relative quantification based on the relative expression of a target gene versus a reference gene. To investigate the physiological changes in gene expression, the relative expression ratio is adequate for the most purposes. (ii) An absolute quantification, based either on an internal or an external calibration curve (1,3). Using such a calibration curve, the methodology has to be highly validated and the identical LightCycler PCR amplification efficiencies for standard material and target cDNA must be confirmed (4–6). Nevertheless, the generation of stable and reliable standard material, either recombinant DNA or recombinant RNA, is very time consuming and it must be precisely quantified (2,7,8). Furthermore, a normalisation of the target gene with an endogenous standard is recommended. Therefore, mainly non-regulated reference genes or housekeeping genes like glyceraldehyde-3-phosphate dehydrogenase (G3PDH or GAPDH), albumin, actins, tubulins, cyclophilin, 18S rRNA or 28S rRNA (9) were applicable. Housekeeping genes are present in all nucleated cell types since they are necessary for basis cell survival. The mRNA synthesis of these genes is considered to be stable and secure in various tissues, even under experimental treatments (9–11). But numerous studies have already shown that the housekeeping genes are regulated and vary under experimental conditions (12–15). To circumvent the high expenditure of design and production of standard material, as well as optimisation and validation of a calibration curve based quantification model, and finally the need for normalisation of the target transcripts to an endogenous housekeeping transcript, a reliable and accurate relative quantification model in real-time RT–PCR is needed.

This study enters into the particular topics of the relative quantification of a target gene in comparison to a reference gene. A new and simple mathematical model for data analysis was established, the application of the new model was tested and compared with available mathematical calculation models. Derived reproducibility, based on intra- and inter-test variation of this relative quantification and accuracy of the model will be discussed.

MATERIALS AND METHODS

RNA source, total RNA extraction and RT

RNA extraction was performed as described previously (16) in bacterial Escherichia coli culture grown either in M9 minimal
media (sample preparation) or LB rich media (control preparation), both with 0.4% glucose concentration (17). RNA integrity was electrophoretically verified by ethidium bromide staining and by OD_{260}/OD_{280} nm absorption ratio >1.95. *Escherichia coli* total RNA (1 \( \mu \)g) was reverse transcribed with 100 U of Superscript II Plus RNase H- Reverse Transcriptase (Gibco BRL Life Technologies, Gaithersburg, MD) using 100 \( \mu \)M random hexamer primers (Pharmacia Biotech, Uppsala, Sweden) according to the manufacturer’s instructions.

**Optimisation of RT–PCR**

Highly purified salt-free primer for target gene 1 (TyrA, tryptophan operon: forward primer, AAG CGT CTG GAA CTG GTT GC; reverse primer, AAA CGC TGT GCG TAA TCG CC), target gene 2 (PyrB, aspartate transcarbamylase: forward primer, GCT CCA ACC AAC ATC CGA; reverse primer, TTC ACG TTG GCG TAC TCG G) and reference gene (Gst, glutathione transferase: forward primer, CTT TGC CGT TAA CCC TAA GGG; reverse primer, GCT GCA ATG TGC TCT AAC CC) were generated (MWG Biotech, Ebersberg, Germany) and optimised to an equal annealing temperature of 60°C. Conditions for all PCRs were optimised in a gradient cycler (Mastercycler Gradient, Eppendorf, Germany) with regard to Taq DNA polymerase (Roche Diagnostics, Basel, Switzerland), forward and reverse primers, MgCl2 concentrations (Roche Diagnostics), dNTP concentrations (Roche Diagnostics) and various annealing temperatures (55–65°C). RT–PCR amplification products were separated on a 4% high resolution NuSieve agarose (FMC Bio Products, Rockland, ME) gel electrophoresis and analysed with the Image Master system (Pharmacia Biotech). Optimised results were transferred on the following LightCycler PCR protocol.

**LightCycler real-time PCR**

For LightCycler reaction a mastermix of the following reaction components was prepared to the indicated end-concentration: 13 \( \mu \)l water, 2.4 \( \mu \)l MgCl2 (4 mM), 0.8 \( \mu \)l forward primer (0.4 \( \mu \)M), 0.8 \( \mu \)l reverse primer (0.4 \( \mu \)M) and 2.0 \( \mu \)l LightCycler (Fast Start DNA Master SYBR Green I; Roche Diagnostics). LightCycler mastermix (19 \( \mu \)l) was filled in the LightCycler glass capillaries and 1 \( \mu \)l cDNA (3.2, 4.0, 4.8, 16, 20 or 24 ng reverse transcribed total RNA) was added as PCR template. Capillaries were closed, centrifuged and placed into the LightCycler rotor. The following LightCycler experimental run protocol was used: denaturation program (95°C for 10 min), amplification and quantification program repeated 40 times (95°C for 15 s, 60°C for 10 s, 72°C for 60 s with a single fluorescence measurement), melting curve program (60–95°C with a heating rate of 0.1°C per second and a continuous fluorescence measurement) and finally a cooling step to 40°C. For the mathematical model it is necessary to determine the crossing points (CP) for each transcript. CP is defined as the point at which the fluorescence rises appreciably above the background fluorescence. ‘Fit Point Method’ must be performed in the LightCycler software 3.3 (Roche Diagnostics), at which CP will be measured at constant fluorescence level (18).

## RESULTS

### Confirmation of primer specificity

Specificity of RT–PCR products was documented with high resolution gel electrophoresis and resulted in a single product with the desired length (TyrA, 978 bp; PyrB, 530 bp; and Gst, Table 1. Intra-assay (test precision) and inter-assay variation (test variability) of LightCycler real-time RT–PCR

<table>
<thead>
<tr>
<th></th>
<th>Intra-assay variation (n = 3)</th>
<th>Inter-assay variation (n = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean CP</td>
<td>CV</td>
</tr>
<tr>
<td>TyrA (sample)</td>
<td>15.020</td>
<td>0.06%</td>
</tr>
<tr>
<td>TyrA (control)</td>
<td>20.303</td>
<td>0.09%</td>
</tr>
<tr>
<td>PyrB (sample)</td>
<td>16.031</td>
<td>2.16%</td>
</tr>
<tr>
<td>PyrB (control)</td>
<td>11.720</td>
<td>1.32%</td>
</tr>
<tr>
<td>Gst (sample)</td>
<td>14.533</td>
<td>0.65%</td>
</tr>
<tr>
<td>Gst (control)</td>
<td>14.277</td>
<td>0.11%</td>
</tr>
</tbody>
</table>

Determination of variation was done in 20 ng reverse transcribed total RNA. Test variation is based on CP variation and expressed as mean CP with CV.

![Figure 1. Determination of real-time PCR efficiencies of reference gene (Gst), target gene 1 (TyrA) and target gene 2 (PyrB). CP cycles versus cDNA (reverse transcribed total RNA) concentration input were plotted to calculate the slope (mean ± SD; n = 3). The corresponding real-time PCR efficiencies were calculated according to the equation: \( E = 10^{-1/\text{slope}} \).](image-url)
transcripts are listed in Table 2. The influence of differing target gene transcript efficiencies were calculated, according to investigated transcripts must be known. Real-time PCR keeping gene transcript. For the calculation of be a stable and secure unregulated transcript, e.g. a house-reference gene transcript. The reference gene could sample of reference gene transcript. The reference gene could 

Real-time PCR efficiencies were calculated from the given slopes in LightCycler software. The corresponding real-time PCR efficiency ($E$) of one cycle in the exponential phase was calculated according to the equation: $E = 10^{\Delta CP_{\text{target}}/\Delta CP_{\text{ref}}}$ (Fig. 1) (18). Investigated transcripts showed high real-time PCR efficiency rates; for TyrA, 2.09; PyrB, 2.16; and Gst, 1.99 in the investigated range from 0.40 to 50 ng cDNA input ($n = 3$) with high linearity (Pearson correlation coefficient $r > 0.95$).

Intra- and inter-assay variation
To confirm accuracy and reproducibility of real-time PCR the intra-assay precision was determined in three repeats within one LightCycler run. Inter-assay variation was investigated in three different experimental runs performed on 3 days using three different premix cups of LightCycler, Fast Start DNA Master SYBR Green I kit (Roche Diagnostics). Determination of variation was done in 20 ng transcribed total RNA (Table 1). Test reproducibility for all investigated transcripts was low in inter-test experiments (<3.91%) and even lower in intra-test experiments (<2.16%). The calculation of test precision and test variability is based on the CP variation from the CP mean value.

Mathematical model for relative quantification in real-time PCR
A new mathematical model was presented to determine the relative quantification of a target gene in comparison to a reference gene. The relative expression ratio ($R$) of a target gene is calculated based on $E$ and the CP deviation of an unknown sample versus a control, and expressed in comparison to a reference gene.

$$\text{ratio} = \frac{(E_{\text{target}})_{\Delta CP_{\text{target}}(\text{control} - \text{sample})}}{(E_{\text{ref}})_{\Delta CP_{\text{ref}}(\text{control} - \text{sample})}}$$

Equation 1 shows a mathematical model of relative expression ratio in real-time PCR. The ratio of a target gene is expressed in a sample versus a control in comparison to a reference gene. $E_{\text{target}}$ is the real-time PCR efficiency of target gene transcript; $E_{\text{ref}}$ is the real-time PCR efficiency of a reference gene transcript; $\Delta CP_{\text{target}}$ is the CP deviation of control – sample of the target gene transcript; $\Delta CP_{\text{ref}}$ = CP deviation of control – sample of reference gene transcript. The reference gene could be a stable and secure unregulated transcript, e.g. a housekeeping gene transcript. For the calculation of $R$, the individual real-time PCR efficiencies and the CD deviation ($\Delta CP$) of the investigated transcripts must be known. Real-time PCR efficiencies were calculated, according to $E = 10^{\Delta CP/\Delta CP}$ (18), as shown in Figure 1. CP deviations of control cDNA minus sample of the target gene and reference genes were calculated according to the derived CP values. Mean CP, variation of CP and $\Delta CP$ values between control and sample of investigated transcripts are listed in Table 2. The influence of differing cDNA input concentrations on $\Delta CP$ are also shown. Intended cDNA input concentration variation of control and sample were compared at different levels (low level, 3.2, 4.0, 4.8 ng cDNA; high level, 16, 20 and 24 ng cDNA). They resulted in stable and constant $\Delta CP$ cycle numbers. In Table 3 the corresponding ratios of target genes in comparison to the reference gene were calculated, through to the established mathematical model (equation 1). The expression ratios of target genes remain stable, even under intended ±20% cDNA variation and low and high cDNA input levels, performed in two runs. A minimal coefficient of variation (CV) of 2.50 and 1.74% was observed, respectively.

Regulation of investigated gene transcripts
All investigated transcript expressions were regulated divergently (Table 3). The expression of Gst was constant, independent of media conditions, and therefore was chosen as endogenous standard or reference gene transcript Fig. 2. TyrA mRNA expression, measured in 20 ng cDNA, was up-regulated 49.1-fold (2.09±2.83) in M9 minimal compared to LB rich medium under high cDNA input conditions. Under the consideration of the reference gene expression the real up-regulation ratio was, on average, 58.5-fold. PyrB mRNA expression was down-regulated under M9 minimal medium conditions by a factor of 27.6 (2.16±3.11). With the normalisation of the endogenous standard transcript, the exact relative expression ratio can be calculated to 23.2.

DISCUSSION
RT followed by PCR is the most powerful tool to amplify small amounts of mRNA (19). Because of its high ramping rates, limited annealing and elongation time, the rapid cycle PCR in the LightCycler system offers stringent reaction conditions to all PCR components and leads to a primer sensitive and template specific PCR (20). The application of fluorescence techniques to real-time PCR combines the PCR amplification, product detection and quantification of newly synthesised DNA, as well as verification in the melting curve analysis. This led to the development of new kinetic RT–PCR
methodologies that are revolutionising the possibilities of mRNA quantification (21).

In this paper, we focused on the relative quantification of target gene transcripts in comparison to a reference gene transcript. A new mathematical model for data analysis was presented to calculate the relative expression ratio on the basis of the PCR efficiency and crossing point deviation of the investigated transcripts (equation 1). The concept of threshold fluorescence is the basis of an accurate and reproducible quantification using fluorescence-based RT–PCR methods (22). Threshold fluorescence is defined as the point at which the fluorescence rises appreciably above the background fluorescence. In the Fit Point Method, the threshold fluorescence and therefore the DNA amount in the capillaries is identical for all samples. CP determination with the ‘Second Derivative Maximum Method’ is not adequate for our mathematical model, because quantification is done at the point of most efficient real-time PCR where the second derivative is at its maximum (18).

A linear relationship between the CP, crossing the threshold fluorescence, and the log of the start molecules input in the reaction is given (18,23). Therefore, quantification will always occur during the exponential phase, and it will not be affected by any reaction components becoming limited in the plateau phase (7). In the established model the relative expression ratio of a target gene is normalised with the expression of an endogenous desirable unregulated reference gene transcript to compensate inter-PCR variations between the runs. The CP of the chosen reference gene is the same in the control and the sample (\(\Delta CP = 0\)). Stable and constant reference gene mRNA levels are given. Under these considerations of an unregulated reference gene transcript, no normalisation is needed and equation 1 can be shortened to equation 2.

\[
\text{ratio} = \frac{E_{\text{target}}}{{E_{\text{ref}}}^0} \frac{\Delta CP_{\text{sample}}}{\Delta CP_{\text{control}}}
\]

Equation 2 shows a mathematical model of relative expression ratio in real-time PCR under constant reference gene expression. CP values in the sample and control are equal and represent ideal housekeeping conditions (\(\Delta CP_{\text{ref}} = 0, E_{\text{ref}} = 1\)).

Two other mathematical models are available for the relative quantification during real-time PCR. The "efficiency calibrated mathematical method for the relative expression ratio in real-time PCR" is presented by Roche Diagnostics in a truncated form in an internal publication (24). The complete equation is,
in principle, the same and the results are in identical relative expression ratio like our model (equation 3).

\[
\frac{\text{ratio}}{(E_{\text{ref}})_{\text{Sample}}} = \frac{(E_{\text{target}})_{\text{Sample}}}{(E_{\text{target}})_{\text{Calibrator}}} - \frac{(E_{\text{ref}})_{\text{Calibrator}}}{}
\]

Efficiency calibrated mathematical method for the relative expression ratio in real-time PCR presented by Soong et al. (24). But the method of calculation in the described mathematical model is hard to understand. The second model available, the ‘Delta–delta method’ for comparing relative expression results between treatments in real-time PCR (equation 4) is presented by PE Applied Biosystems (Perkin Elmer, Forster City, CA).

\[
\text{ratio} = 2^{\Delta\Delta CP}
\]

Equation 4 shows a mathematical delta–delta method for comparing relative expression results between treatments in real-time PCR developed by PE Applied Biosystems (Perkin Elmer). Optimal and identical real-time amplification efficiencies of target and reference gene of PE Applied Biosystem (Perkin Elmer). Optimal and identical real-time amplification efficiencies of target and reference similarly. All components, like enzyme inhibitors, and cDNA synthesis background interferences retrieved from extracted tissue step RT–PCR, using random hexamer primers, all possible interferences during RT will influence all target transcripts as well as the internal reference transcript in parallel. Occurring background interferences retrieved from extracted tissue components, like enzyme inhibitors, and cDNA synthesis efficiency were related to target and reference similarly. All products underwent identical reaction conditions during RT and variations only disappear during real-time PCR. Any source of error during RT will be compensated through the model itself. Widely distributed single-step RT–PCR models are not applicable, because in each reaction set-up and for each investigated factor individual and slightly different RT conditions will occur. Therefore, the variation in a two-step RT–PCR will always be lower, and the reproducibility of the assay will be higher, that in a single-step RT–PCR (8). Reproducibility of the developed mathematical model was dependent on the exact determination of real-time amplification efficiencies and on the given low LightCycler CP variability. In our mathematical model the necessary reliability and reproducibility was given, which was confirmed by high accuracy and a relative error of <2.5% using low and high template concentration input.

### CONCLUSION

LightCycler real-time PCR using SYBR Green I fluorescence dye is a rapid and sensitive method to detect low amounts of mRNA molecules and therefore offers important physiological insights on mRNA expression level. The established mathematical model is presented in order to better understand the mode of analysis in relative quantification in real-time RT–PCR. It is only dependent on ΔCP and amplification efficiency of the transcripts. No additional artificial nucleic acids, like recombinant nucleic acid constructs in external calibration curve models, are needed. Reproducibility of LightCycler RT–PCR in general and the minimal error rate of the model allows for an accurate determination of the relative expression ratio. Even different cDNA input resulted in minor variations. Relative expression is adequate for the most relevant physiological expression changes. In future it is not necessary to establish more complex and time consuming quantification models based on calibration curves. For the differential display of mRNA the relative expression ratio is an ideal and simple tool for the verification of RNA or DNA array chip technology results.

### ACKNOWLEDGEMENTS

The author thanks D.Schmidt for technical assistance. Primers, primer sequences and samples were kindly donated by Drs...
S. Wegener and W. Mann in collaboration with the BioChip division of MWG Biotech in Ebersberg, Germany.

REFERENCES


