Lehrstuhl für Physiologie
Fakultät Wissenschaftszentrum Weihenstephan
Technische Universität München

Quantitative real-time RT-PCR based transcriptomics: Improvement of evaluation methods

Aleš Tichopád

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung
des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. L. Dempfle
Prüfer der Dissertation:	1. Priv.-Doz. Dr. M. W. Pfaffl
	2. Univ.-Prof. Dr. J. Polster

Die Dissertation wurde am 29.06.2004 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 16.08.2004 angenommen.

Table of content

ZUSAMMENFASSUNG

Die quantitative Reverse-Transkription mit Polymerase-Kettenreaktion (qRT-PCR) ist eine neue Methode, um geringste mRNA Mengen in biologischen Proben zuverlässig zu bestimmen. Da die Stärke der Fluoreszenz des Reporterfarbstoffes proportional zur gebildeten DNA-Menge sein sollte, ermöglicht die Aufnahme der Fluoreszenz eine Darstellung des gesamten Reaktionsverlaufs. Demnach kann aus diesem Reaktionsverlauf ein Rückschluss auf die eigentliche Ausgangskonzentration der mRNA gezogen werden.

Während der RNA-Extraktion können hemmende Substanzen mit isoliert werden, die im Folgenden die Reverse-Transkription (RT) als auch die PCR inhibieren und somit zu einem probenspezifischen Verlauf der Reaktion führen. Zusätzlich variiert die PCR-Effizienz nicht nur zwischen zwei Proben, sondern auch während der Registrierung der PCR-Kurve einer einzelnen Probe. Aus diesem Grund ist die korrekte Bestimmung der PCR-Effizienz jeder einzelnen Probe wie auch eine zweckmäßige Standardisierung der Expressionsrohwerte eine wichtige Voraussetzung für die korrekte Interpretation der Ergebnisse.

Um eine Lösung dieser Probleme zu finden, wurde eine Reihe von biologischen Versuchen durchgeführt, in denen die RNA aus verschiedenen Geweben von Schaf und Rind, so wie auch aus Zellkultur-Leukozyten extrahiert wurde. Eine konstante Menge der RNA wurde dann in die cDNA übersetzt. Alle PCR-Läufe wurden mittels eines LightCyclers durchgeführt und die Fluoreszenzrohdaten direkt von der LightCycler Software gespeichert.

Mit Hilfe dieser biologischen Daten wurden mathematische Modelle sowie statistische Verfahren entwickelt und validiert, mit denen man den optimalen Quantifizierungsbereich ermitteln, die Effizienz der real-time PCR erstmals exakt

bestimmen, die Heterogenität zwischen experimentellen Proben untersuchen und somit die Qualität der Expressionsergebnisse verfeinern kann. Aufbauend auf diese Standardisierungen, wurde ein Entscheidungsalgorithmus für die zweckmäßige Auswertung der qRT-PCR Daten entwickelt.

SUMMARY

Quantitative real-time polymerase chain reaction (qRT-PCR) is a new method for reliable quantification of low-abundance mRNA in biological samples. Since the strength of the fluorescence signal emitted by the report dye is proportional to the produced DNA amount, the fluorescence monitoring enables visualisation of the full reaction trajectory. The reaction trajectory can be then extrapolated back to an input concentration.

RNA extraction can introduce unwanted contaminants into the sample, inhibiting the reverse transcription (RT) as well as the PCR reaction. These inhibitions cause then the reaction to precede sample-specific. In addition, the amplification efficiency varies not only between samples, but also along the recorded amplification trajectory of a single sample. Consequently, a correct determination of each probe's PCR efficiency as well as a good standardization of the raw expression estimators is of great importance for a correct interpretation of results.

To find a solution to above problems a series of biological experiments with RNA extracted from various ovine and bovine tissues and from cultured leukocytes was carried out. Constant amount of RNA was then reverse–transcribed to cDNA. All PCR runs were performed on a LightCycler instrument and Fluorescence data was saved in the LightCycler software.

Based on this data, mathematical models together with statistical procedures were developed and validated. These can investigate the optimal quantification range and exactly determine its real-time PCR efficiency. Additionally, methods were developed to disclose heterogeneity between probes. All these procedures contribute to better quality of results obtained. Resulting from these standardisations, a decision algorithm for a proper analysis of the qRT-PCR data was designed.

ABBREVIATIONS

AMV	*Avian Myeloblastosis Virus*
ANOVA	*analysis of variance*
cDNA	*complementary DNA*
CP	*crossing-point (raw PCR quantification unit)*
Ct	*threshold cycle (raw PCR quantification unit)*
DNA	*deoxyribonucleic acid*
dsDNA	*double-stranded DNA*
E	*real amplification efficiency*
ε	*reported amplification efficiency*
GAPDH	*Glyceraldehyd-3-Phosphate Dehydrogenase*
IL-6	*interleukin 6*
KOD	*kinetic outlier detection*
LPS	*Lipopolysacharid*
MMLV	*Malloney murine leukemia virus*
MMLV H⁻	*Malloney murine leukemia virus reverse transcriptase RNase H-minus*
mRNA	*messenger RNA*
PCR	*polymerase chain reaction*
PrP^c	*prion protein*
qRT-PCR	*quantitative Reverse-Transkription mit Polymerase-Kettenreaktion*
recDNA	*recombinant DNA*
recRNA	*recombinant RNA*
RNA	*ribonucleic acid*
rRNA	*ribosomal RNA*
RT	*reverse transcription*
RT-PCR	*reverse transcription polymerase chain reaction*
ssDNA	*single-stranded DNA*
WBC	*while blood cells*
18S	*18 S rRNA*
28S	*28 S rRNA*

1 INTRODUCTION

Molecular diagnostics is a fast developing discipline of the economically important field of biotechnologies. It has found its place in biological, agricultural, veterinary, medical, and forensic science and praxis. Polymerase chain reaction (PCR) based technologies have been established in most of laboratories involved in biomedical science. Bright spectrum of modified applications of the PCR serves to a detection of pathogen organisms in food and environment (Starnbach et al., 1989; Wilson et al., 1993). Also the application of PCR in gene-expression studies remains an example of rapidly innovating field (Orlando et al., 1998; Meuer et al., 2001).

So far, the real-time PCR following reverse-transcription reaction (RT), real-time RT-PCR (Larrick, 1992; Ferré, 1992), is the leading approach adopted in quantifications of low-copy transcripts (Bustin, 2000; Schmittgen, 2001; Klein, 2002). The fact that several nucleic acid molecules can be amplified up to micrograms opens the possibility to study gene regulation even in a single cell (Liss, 2002). The recent introduction of myriad of fluorescence monitoring techniques into PCR allowed documentation of the amplification process in the "real-time" fashion (Holland et al., 1991; Higuchi et al., 1993; Morrison et al., 1998; Marras et al., 1999; Whitcombe et al., 1999; de Silva and Wittwer, 2000). Several real-time PCR platforms are manufactured, of those the *LightCycler* (Roche Diagnostics, Switzerland; Wittwer et al., 1997; Rasmussen, 2001), *ABI Prism* (Applied Biosystemes, USA; Livak, 2001), *Rotor-Gene* (Corbett Research, Australia; User bulletin available under: http://www.corbettresearch.com/rg3000web.pdf) and *iCycler* (BioRad, USA; Cunnick and Jiang, Bio-Rad bulletin 2806) systems are of the widest use.

In this thesis, further improvements of evaluation methods for the real-time RT-PCR data (Muller et al., 2002) obtained mostly on the LightCycler are presented. A

compact series of evaluation steps is suggested that helps investigator to minimise error generated during the PCR procedure and data evaluation.

1.1 Sample preparation, RNA extraction and cDNA synthesis

The subspecies of mRNA transcribed from the gene of interest is the carrier of information about its expression intensity. Therefore the first step on the way to obtain a reliable information about gene-expression intensity is the extraction of either the total RNA or an mRNA from biological sample. The purity and integrity of the extract is the fundamental prerequisite of a good procedure (Swift et al., 2000). It must be free of any DNA contamination, without any potential PCR inhibitors, and well preserved from degradation by RNAses. Phenol, salts or ethanol are known contaminants affecting the PCR performance (Wilson, 1997; Freeman et al., 1999). Up to now, however, no perfect extraction method is available, and the RNA extracted always contains some DNA and proteins (Mannhalter et al. 2000). It was shown that some residual contaminants remain in the sample after RNA extraction (Rossen et al., 1992; Wilson et al., 1993; Wilson, 1997). These tissue-specific residua can inhibit the RT as well as the PCR (Tichopad et al. 2004). In addition, the efficacy of the RNA extraction is, similarly, tissue dependent (Mannhalter et al. 2000). This can cause a systematic error where the gene-expression is expressed per weight of tissue.

The sampling storage and extraction can have a devastating affect on the sample if done wrong. Different times between the slaughtering and the cold storage of samples can be responsible for discrepancies between samples due to post-mortal changes such as RNA cleavage. Possibly fast sampling followed by immediate freezing in liquid nitrogen is perhaps the most often used and most effective method.

RNA molecules cannot serve as a template for the PCR. Therefore, to acquire information about particular gene's expression, the total RNA or mRNA must be reverse-transcribed into its complementary DNA copy in the reverse transcription reaction. The resulting cDNA is single-stranded (ssDNA). The Moloney Murine Leukemia Virus (MMLV) reverse transcriptase is the enzyme of choice to reverse transcribe the RNA into cDNA. Currently, the MMLV H^- is the most frequently used modified version of the MMLV enzyme (Wong et al., 1998). This enzyme facilitates synthesis of a full-length RNA molecules with a high fidelity since its RNAse activity (i.e. RNA digesting exonuclease activity) is significantly lower than in alternative Avian Myeloblastosis Virus (AMV) reverse transcriptase (de Stefano et al., 1991). The RT step can be primed with specific primers, oligo-dT, random hexamer, octamer or decamer primers. The right choice demands a careful consideration. The specific primers decrease background priming of non-specific sequences, whereas the random primers and the oligo-dT primers maximise the chance of successful RT reaction in low RNA samples (Schwabe, 2000). These methods are also less discriminatory to various sequences quantified within one study, and thus facilitate a better comparison between these sequences.

Sample preparation, RNA extraction, and the RT reaction is believed to be a significant source of variation in the entire assay. Precautions are therefore of great importance.

1.2 Principle of the quantitative real-time PCR

The PCR is a chain reaction progressing in a doubling fashion. That is, each selected DNA molecule becomes a target template for synthesis of its one new complementary copy within one cycle of the polymerase reaction. This reaction is facilitated by DNA

polymerase enzyme and primed by sequence specific primer pair. Pair of selected primers flanks a sequence of DNA from both sides, each primer on one side of the two complementary DNA strands. If just a single DNA strand is used as a target template, the primer complementary to this strand anneals and initiates the synthesis of complementary strand in the 5' to 3' direction, proceeding until the end of the DNA strand is reached or the synthesis exhausted (Kainz, 2000). The newly synthesised strand can then serve as a template for synthesis of the complementary strand primed by the second primer.

The full procedure of reaction is facilitated by precisely set temperature regimen. This regimen repeats and the polymerase reaction is initiated again and again, producing two fold higher template concentrations in each new PCR cycle. Such an ideal doubling fashion of the PCR reaction can be described by the following model:

$$P=T\cdot 2^n \qquad [1]$$

In this model, ***P*** is the product measured after ***n*** cycles and ***T*** is the starting amount of the target sequence. The doubling fashion is given by the "2" in the base and represents the ideal amplification performance of the PCR.

The PCR reaction can, nevertheless, proceed in a fashion different from the perfect doubling. Following amplification model takes variable amplification efficiency into account:

$$P=T(1+E)^n \qquad [2]$$

where ***P*** is the product measured after ***n*** cycles, ***T*** is the starting amount of the target sequence, ***E*** is the *real amplification efficiency* (Souazé et al., 1996; Peccaud and Jacob, 1998; Tichopad et al., 2003a) expressed as the percentage of target molecules copied in one average PCR cycle (from 0, representing no amplification, to 1, representing the ideal doubling).

The course of the reaction can be visualised by the emitted fluorescence (Higuchi et al., 1993; Morrison et al., 1998; de Silva and Wittwer, 2000). The real-time PCR detection system monitors rather the fluorescence emitted in each cycle ***n*** then the concentration of molecules itself. Therefore the ***P*** from the equations [1 and 2] should be replaced by ***f*** denoting the fluorescence measured and the ***T*** should be replaced by α denoting the fluorescence emitted by the starting nucleic acid input. The system without any nucleic acid input, however, also emits some background fluorescence, here γ_0.

Then the equation [2] can be written as:

$$f = \gamma_0 + \alpha\varepsilon^n \qquad [3]$$

The ε is the *reported amplification efficiency* in the exponential phase of the real-time PCR with values between 1 and 2 (2 for ideal doubling). If the fluorescence monitoring reflects the amplification of target sequence truly, it holds that $\varepsilon = 1 + E$.

In each cycle of the PCR a higher amount of fluorescence is emitted than in the previous cycle, provided, the polymerase reaction was, at least minimally, successful. According to this principle, reaction with higher starting concentration of template must reach any chosen fluorescence threshold level sooner than any reaction with lower concentration of the same template (Gibson et al. 1996; figure 1). The number

of cycles necessary to reach the selected threshold fluorescence is the fundamental quantitative unit of the real-time PCR assay called crossing point – CP[1] (Rasmussen, 2001), or threshold cycle – C_t (Higuchi et al., 1993). Therefore, if two samples – *sample 1* and *sample 2* – are real-time PCR assayed the difference between their CPs in the exponent reports about the ratio between the initial concentrations of *sample 1* and *sample 2*. The ratio *R* can be then calculated as follows:

$$R = 2^{\Delta CP(sample1-sample2)} \qquad [4]$$

or generally for varying amplification efficiency ε

$$R = \varepsilon^{\Delta CP(sample1-sample2)} \qquad [5]$$

First, the ε must be estimated to utilise the equation [5] for quantification. In addition, ε must be the same in both samples for the equation [5] to be correct. This computational model can be applied on studies of *pre-* vs. *post-treatment* type where ε is supposed to remain constant. The *R* than says how many times the *post-treated* gene-expression is down- or up-regulated compared to its pre-treated state.

[1] The CP term will be used throughout the following text as it is adopted within the LightCycler PCR platform familiar to author.

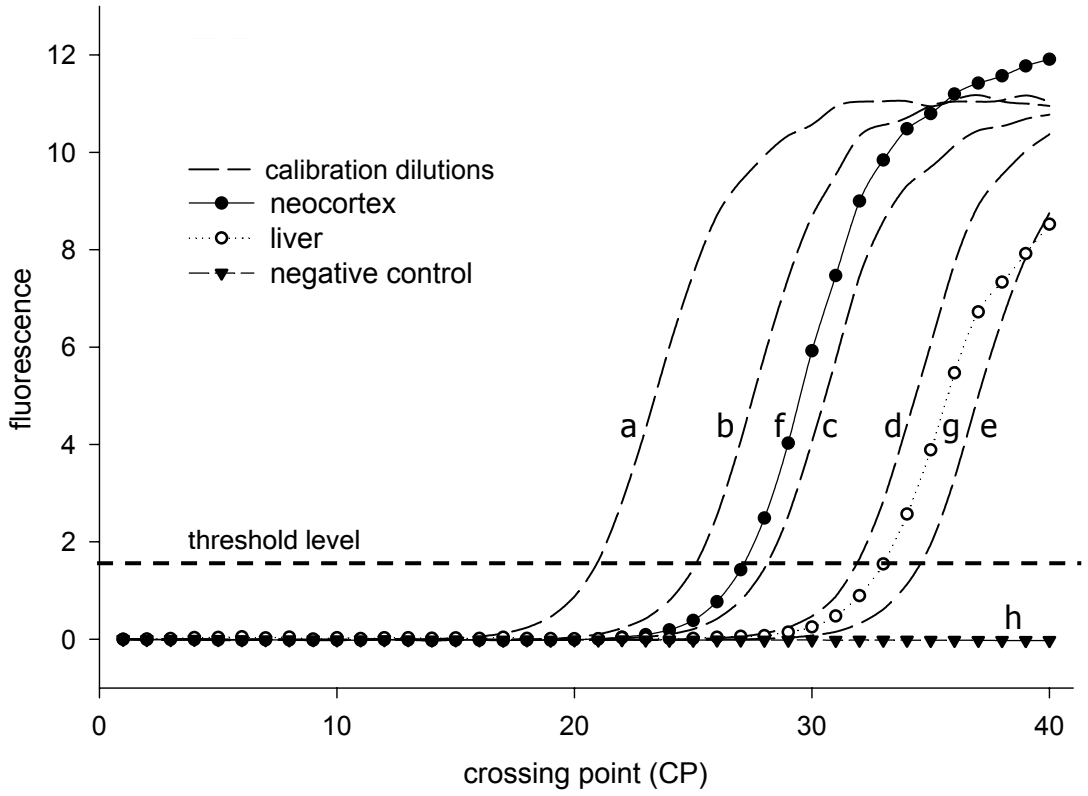


Figure 1. Example of real-time PCR amplification curves obtained by plotting fluorescence data against their cycle number (amplification of 262 bp recDNA and biological fragment of bovine prion sequence). Five calibration dilutions (from left to right: a) $2x10^7$ copies, b) $2x10^6$ copies, c) $2x10^5$ copies, d) $2x10^4$ copies, and e) $2x10^3$ copies) are shown together with two biological sample of bovine neocortex f) $4.25x10^5$ copies, CP=26.87 and liver g) $7.41x10^3$ copies, CP=32.50, and a negative control h) – without nucleic acid input. The subjectively set threshold level is marked with the dashed line. The calibration dilutions produced following CPs: from left to right: a) 20.24, b) 24.96, c) 28.32, d) 31.91, and e) 34.62. Both biological samples produced following CPs: $CP_{neocortex}$ =27.17 and CP_{liver} =32.74. The CP values were obtained according to Rasmussen (2001; see later in 1.2.2).

1.2.1 Description of the PCR kinetics

The amplification of nucleic acids according to the equation [2] can be described as an exponential growth. This fashion of the amplification is, however, not present during the whole course of the reaction, but only in its initial part. As the reaction

proceeds, the reaction's exponential character decays. The smoothed kinetics of the reaction gives typical S-like amplification curve (Schnell and Mendoza, 1997; Liu and Saint, 2002b; Tichopad et al., 2003a). Three portions can be distinguished within the amplification curve (figure 2):

Portion 1 – Background: Although the exponential character of the template amplification is present, it cannot be detected as it lies under a detection threshold of the PCR platform. The amplification is nevertheless present, and proceeds according to the equation [2].

Portion 2 – Growth: This portion of the increasing fluorescence acquisition can be subdivided into two phases:

First phase poses strong exponential growth trend of template amplification that is already above the detection threshold. The fluorescence is already monitored in this phase, and the amplification can be thus described by equation [3].

The following second phase above the exponential also poses some growth trend. This growth is, however, no longer exponential as the fluorescence observations start diverging from the exponential trend (figure 3). This phase can no longer be described by the equation [3].

There is a hardly detectable smooth transition between these both phases. Determination of the end of the exponential phase becomes an interesting challenge of a crucial importance for correct amplification efficiency estimation (Peccoud and Jacob, 1998; Liu and Saint, 2000a; Tichopad et al. 2003a).

Portion 3 – Plateau. In this phase of the fluorescence acquisition, the reaction becomes exhausted (Kainz, 2000). Its growth trend decays and its course turns more stochastic (Peccaud and Jacob, 1996) due to non-specific products generated. The state of the reaction in this phase bears no correlation to the initial template

concentration in the reaction mix. The post PCR template-non-specific products inflate the quantitative information contained in the fluorescence signal. This portion is no longer suited for measurement purposes.

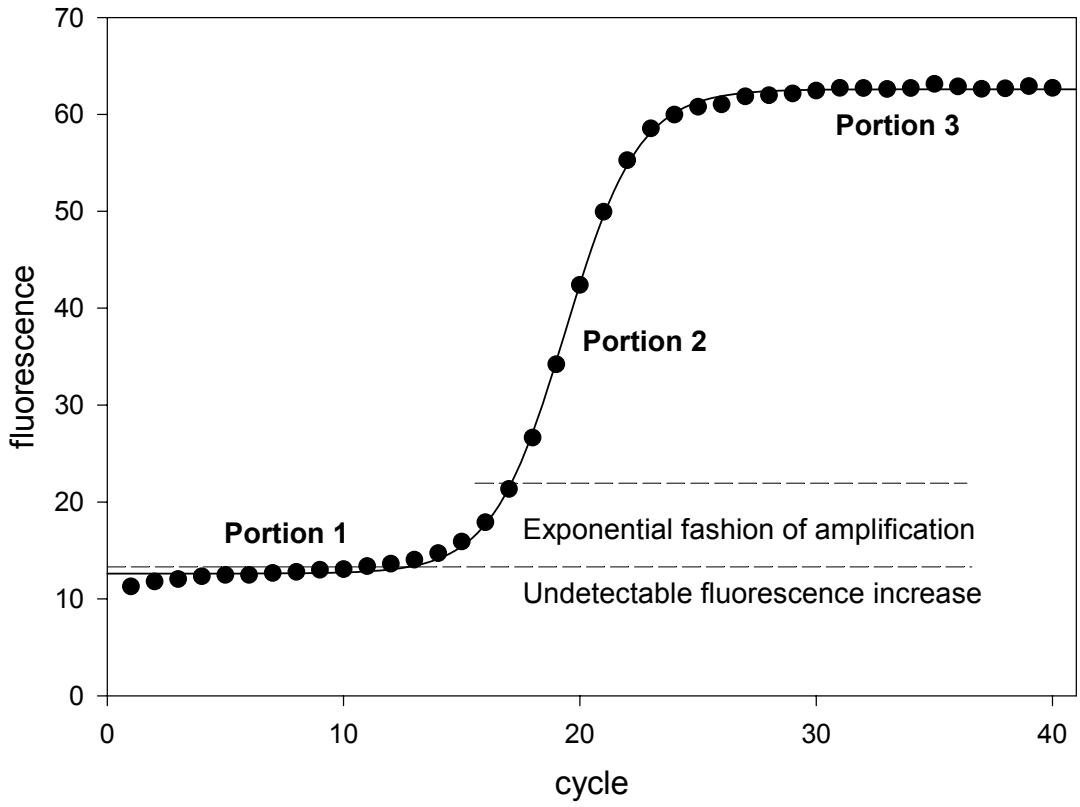


Figure 2. Portions of the amplification curve representing course of the PCR reaction (amplification of 262 bp recDNA fragment of bovine PrP^c). Fluorescence detection threshold is indicated by the lower dashed line and terminates the portion 1 of the amplification curve. Under this threshold the increase of fluorescence signal due to amplification gets lost in the background fluorescence noise. The exponential fashion of amplification can be first significantly detected above this threshold and usually takes no more than ten cycles (till the upper dashed line). Further in the portion 2, the exponential fashion of amplification decays with a long smooth transition into plateau phase. Finally, the reaction enters the plateau phase in the portion 3.

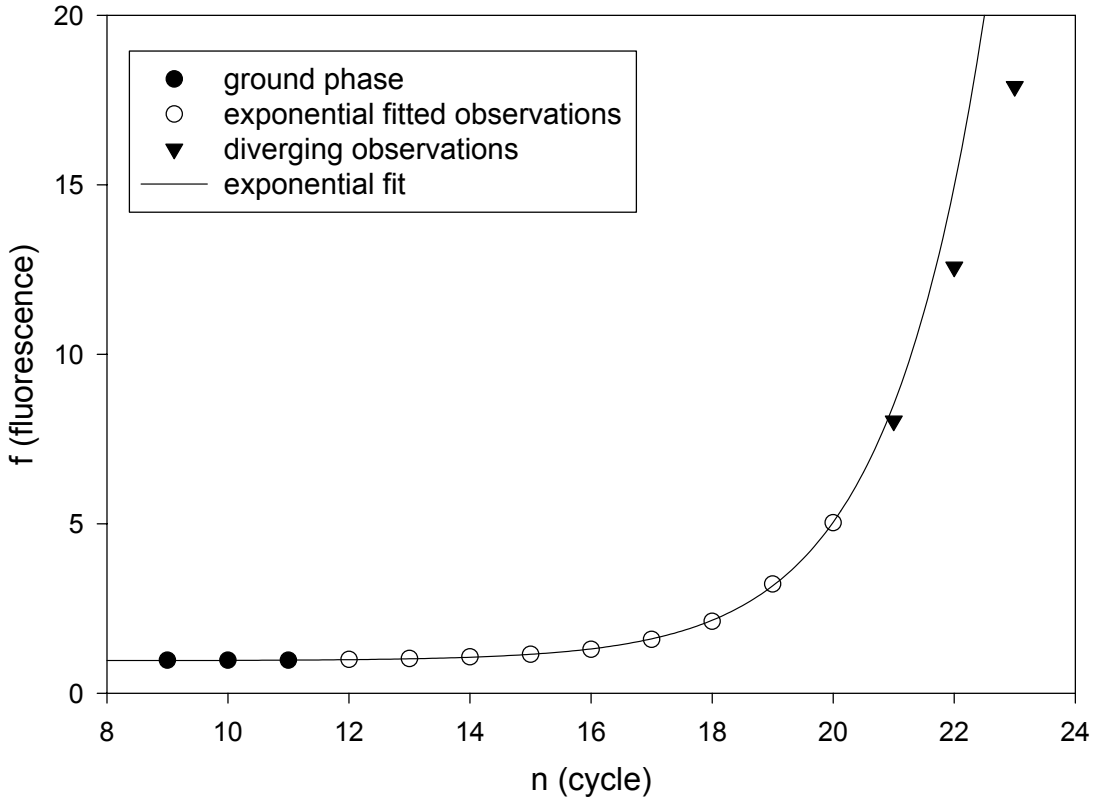


Figure 3. Decay of the exponential trend of the template amplification by PCR (amplification of 197 bp sequence of bovine GAPDH in biological sample from muscle). The exponential model $f = \gamma_0 + \alpha\varepsilon^n$ was fitted over the fluorescence data. In this model *f* denotes the measured fluorescence, γ_0 is the background fluorescence, α denotes the fluorescence emitted by the starting nucleic acid input, ε is the reported amplification efficiency with value between 1 and 2, and *n* is the cycle number.

1.2.2 The crossing point and the quantification event

The quantification of a starting target sequence concentration in the sample is conducted in form of fluorescence acquisition at determined fluorescence threshold. The fractional number of cycles at this point, crossing point, is the measure of the starting target sequence concentration. To take use of most of the quantitative power of the real-time PCR, the crossing point acquisition must take place as soon as

possible (Tichopad et al. 2003a). In praxis it means, to get rid of the uninformative background phase and to perform the quantification step within the exponential phase of the second portion of the amplification curve. Therefore, as soon as it is obvious that the amplification course entered the detectable exponential phase the reaction can be theoretically terminated, because the information necessary for the quantitative judgement has been gathered. Further course of the amplification curve is, from this point of view, no longer interesting. This is the main advantage of the real-time PCR in contrary to older *end-point* method. In the *end-point* method no real-time monitoring is employed and the reaction is terminated up to investigator's subjective assumption (Freeman et al. 1999; Schmittgen et al., 2000). Such a termination can be, however, done too late within the plateau phase where the reaction is no longer suited for the quantification purposes.

In the real-time PCR, the crossing point acquisition takes place long before the reaction reaches its plateau. In such an early portion of PCR trajectory the amplification has really the assumed exponential character as described by equation [2]. In the early exponential portion of amplification kinetics the threshold fluorescence is set (Rasmussen, 2001).

There are two ways of threshold level setting: It can be done arbitrary, setting the threshold into a portion of the kinetic curve that is subjectively considered exponential (Rasmussen, 2001). Alternatively, the threshold fluorescence value can be obtained by applying some computing algorithm. In this case, there can be an individual threshold for each amplification curve. As an example, the maximum of the second or, generally, n^{th} derivative of the smoothed amplification kinetics gives a good and justified crossing point (Wittwer et al., 1999). The great advantage of this method is, that it is not affected by the individual decision on the best threshold fluorescence

value (Rasmussen, 2001). The n^{th} derivative maximum is in constant exact relation to the staring template concentration.

1.2.3 Smoothing of the PCR amplification by empirical model

As the PCR kinetics is not fully understood up to now, there is no satisfactory theoretical mathematical model describing the reaction from its beginning to the end in its plateau phase. The system of saturation and inhibition resulting in the exhaustion of the reaction is very complicated (Schnell and Mendoza, 1997; Kainz, 2000). Therefore the classic model of exponential PCR as given by equation [2] and [3] applies only to a certain initial portion of the entire kinetic trajectory. As a partial solution, several empirical non-linear models can be constructed to smooth the full amplification trajectory. The *four–parametric sigmoid model*

$$f = y_0 + \frac{a}{1 + e^{-(\frac{n-n_0}{b})}} \qquad [4]$$

can be fitted over all amplification fluorescence data, often with a determination $r^2 > 0.999$ (Tichopad et al., 2002; Tichopad et all., 2004). In this model, $\boldsymbol{f}$ is the value of function computed (fluorescence after $\boldsymbol{n}$ cycles), $\boldsymbol{y_0}$ is the ground fluorescence, $\boldsymbol{a}$ is the difference between maximal fluorescence acquired and the ground fluorescence, $\boldsymbol{e}$ is the natural logarithm base, $\boldsymbol{n}$ is the actual cycle number, $\boldsymbol{n_0}$ is the first derivative maximum of the function or the inflexion point of the curve, and $\boldsymbol{b}$ describes the slope of curve in $\boldsymbol{n_0}$ (figure 4).

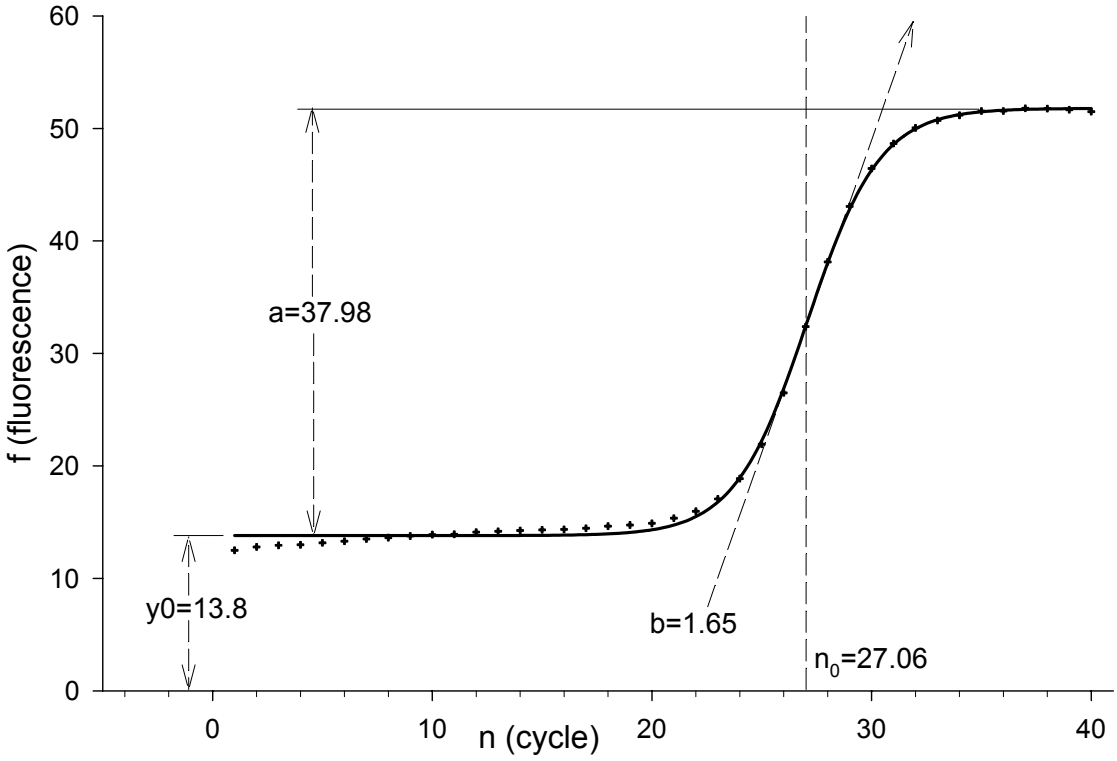


Figure 4. Four-parametric sigmoid model. This model is defined as $f = y_0 + \frac{a}{1 + e^{-(\frac{n-n_0}{b})}}$.

One fluorescence data set from the amplification of 234 bp sequence of ovine β-actin was taken as an example. In this model, y_0 is the ground fluorescence, a is the difference between maximal fluorescence acquired in the run and the ground fluorescence, n_0 is the first derivative maximum of the function or the inflexion point of the curve, and b describes the slope of curve in n_0.

The following general scheme can be given:

a high ***a*** corresponds high PCR product obtained after all cycles,

b low ***b*** corresponds high reaction efficiency,

n_0 high n_0 corresponds delay of the reaction performance due to low input concentration or high ***b*** parameter. It is comparable to ***CP***,

y_0 high y_0 detects high background fluorescence.

Surely, there are some common components behind this parameters. Some sophisticated statistical analysis would therefore be of help to disclose main components behind them.

The second derivative maximum can be computed from this model as an alternative to the CP value generated by the real-time PCR platform. The computation is as follows, beginning from the equation [4]:

$$f'(n) = \frac{a}{b} \cdot \frac{e^{-\frac{n-n_0}{b}}}{\left(1+e^{-\frac{n-n_0}{b}}\right)^2} \qquad [5]$$

$$f''(n) = \frac{a}{b^2} \cdot \frac{e^{-\frac{n-n_0}{b}} - e^{-2\frac{n-n_0}{b}}}{\left(1+e^{-\frac{n-n_0}{b}}\right)^3} \qquad [6]$$

$$f'''(n) = \frac{a}{b^3} \cdot \frac{e^{-\frac{n-n_0}{b}} - 4e^{-2\frac{n-n_0}{b}} + e^{-3\frac{n-n_0}{b}}}{\left(1+e^{-\frac{n-n_0}{b}}\right)^4} \qquad [7]$$

The first [equation 5], the second [equation 6], and the third derivation [equation 7] of the model [equation 4] are calculated. To result in a second derivative maximum the third derivation has to be null: f’’’(n) = 0. Two second derivative maximums are given for $n \approx n_0 \pm 1.317 \cdot b$, whereas only the first “positive maximum” is relevant for an approximation of the CP. Therefore,

$$CP \Rightarrow n = n_0 - 1.317 \cdot b \qquad [8]$$

Also another model was applied to smooth the whole amplification course of the reaction (Tichopad et al, 2003a). The *four-parametric logistic model*

$$f(x) = y_0 + \frac{a}{1 + \left(\frac{n}{n_0}\right)^b} \qquad [9]$$

generates S-like curve, and can fit the fluorescence data with a comparable goodness. The four parameters denote the same geometric properties of the smoothing curve as in the equation [4]. The scaling of the parameter ***b*** is, however, different from the four-parametric sigmoid model given by the equation [4]. This model shows no central-point-symmetry and is therefore more flexible.

1.3 Quantification strategies

The quantification of gene-expression is never done in a single sample. Such a result, actually only a fluorescence value, would be irrelevant, saying nothing about regulatory biological process in the organism. Quantification in several, but at least two, samples must always be carried out to extract minimal useful information. But also such a result is just very simplified and almost poor of any biological relevance. This is because the error factors linked to each sample are different and the comparison is therefore not possible without a great deal of assumptions. Two main methods of quantification of gene-expression data are available:

1.3.1 Absolute quantification

The crossing point of studied sample is confronted with a calibration curve constructed on known concentrations of the same target sequence. The result obtained is a number of transcript copies in the sample which can be then recalculated and expressed per g tissue, ng total extracted RNA, one cell or another denominator

(Pfaffl et al., 2001b; Tichopad et al., 2003b). As the method quantifies the absolute amount of transcript it is called the absolute quantification (Bustin, 2000). The right choice of denominator depends on the questions asked, and also affects the quality of results (Ferré, 1992). The chosen denominator always reflects just a distinct part of the whole assay. For example, if the quantification data are expressed per amount of total RNA extracted, the influence of RNA subpopulations (e.g. transfer RNA or ribosomal RNA) on the total amount of the RNA extracted will be omitted.

The calibration curve is constructed on recombinant DNA (recDNA) or RNA (recRNA) (Pfaffl and Hageleit, 2001). Also synthetic nucleotide or product of previous PCR can be used. The method performs well as long as a proper range of dilution is chosen. The loos of linearity at the beginning and at the end of the dilution curve are often discussed problems (Hocquette and Brandstetter, 2002). In Tichopad et al. (2002) non-pathogen prion protein, PrP^c (Prussiner, 1998), was quantified using absolute quantification with the calibration curve constructed out of five dilutions. The figure (1) shows the position of the amplification curves of two biological samples; neocortex and liver within range given by five calibration samples.

1.3.2 Relative quantification

If just known dilutions from the steady-state transcript are used as the calibration curve, the result has a form of relative up-/down-regulation (Bustin, 2000) from the steady-state. As the steady-state the healthy state or the state before experiment is understood. The change from the steady-state during any experimental treatment or pathological change is then the studied goal. The method is then called ‘the relative quantification’ (Livak and Schmittgen, 2001), and the steady-state sample serves as a control sample. For example, a result obtained in this way reports that the IL-6

expression in cultured white blood cells (WBC) confronted with Lipopolysacharid (LPS) endotoxin is four fold up-regulated in comparison to its control (i.e. cells before endotoxin injection). The relation between concentration of the studied sequence in the sample and the known steady-state control concentration is described by equations [5] and can be rewritten as

$$R = \varepsilon^{\Delta CP\ (control\text{-}sample)} \qquad [10]$$

where R is the ratio between gene-expression in the control and studied sample. Since there is only one parameter for amplification efficiency, ε, in the equation [10], it is assumed that the assay shows a homogeneous performance for both control and sample.

Alternatively, if a heterogeneous performance between sample and control is expected, the assay must be standardised. To standardise for different assay's performances, some other gene sequence is quantified together with the studied sequence either simultaneously in the same sample or in a parallel fashion (Serazin-Leroy et al., 1998; Suzuki et al., 2000). The standardisation with a reference gene whose expression is believed to be constant, housekeeping gene (Warrington et al., 2000), is useful where the some disturbance during extraction, RT reaction, storage, and PCR itself can introduce some discrepancy between samples. The sequence of the standard gene is present in the sample together with the target sequence during the whole assay, and mimics the target sequence as to all errors and disturbances during the assay. Many genes such as tubulins, actins Glyceraldehyd-3-Phosphate Dehydrogenase (GAPDH), albumins, cyclophilin, micro-globulins 18S or 28S rRNA have been described in literature whose expression is believed to remain constant

under an experimental intervention. On the other hand, some of these genes were also reported to undergo regulation under defined conditions (Chang et al., 1998; Foss et al., 1998; Thellin et al., 1999; Schmittgen and Zakrajsek, 2000). Several relative expression models have been suggested up to now. Standardisation model $R = 2^{-\Delta\Delta CP}$ (Livak and Schmittgen, 2001) can be applied where the same amplification efficiency in target sequence and the standard gene is assumed.

More recently, model including amplification efficiency correction have been shown by Pfaffl (2001a):

$$R = \frac{\varepsilon_{target}^{\Delta CP_{target}(control-sample)}}{\varepsilon_{standard}^{\Delta CP_{standard}(control-sample)}} \qquad [11]$$

were ***R*** denotes the standardised computed expression ratio between control target gene and studied sample target gene. The ε denotes the amplification efficiency and the Δ***CP$_{target}$(control-sample)*** or Δ***CP$_{standard}$(control-sample)*** is the difference between CP value of the control and the CP value of the studied sample. If the control sample is taken before experiment and the studied sample after it then a result of R=0.5 says that the experimental treatment caused two fold down-regulation.

1.4 Amplification efficiency correction

The fundamental parameter of the PCR reaction performance is its *real amplification efficiency* ***E*** from equation [2] (Peccaud and Jacob, 1998; Rasmussen, 2001; Liu and Saint, 2002a; Tichopad et al. 2003a). It can be also understood as a chance between 0% and 100%, that a single template molecule will get replicated in the following PCR cycle. If the reaction conditions are optimal, the chance for a molecule to be

successfully replicated is high and such a reaction performs well. This parameter can be estimated from the acquired fluorescence signal in form of the *reported amplification efficiency* ε. It should hold that $\varepsilon = 1 + E$ where the increase of fluorescence signal reflects the increase of template concentration tightly. Even if two samples have exactly the same starting concentration of the target sequence, any difference in the amplification efficiency would result in different quantitative results (Tichopad et al., 2002; Tichopad et al. 2003a). If the ε_{sample} = 0.8 and $\varepsilon_{control}$ = 0.9 then the ***R*** will be approximately 3.6 fold underestimated after 25 cycles if calculated according to equation [10]. This is a direct consequence of the exponential amplification of the initial error. The exact estimation of ε is crucial where discrepancies in performance between samples are expected. If ε is known, a compensation algorithm can be applied (Livak and Meijerlink et al., 2001; Pfaffl, 2001a; Schmittgen, 2001; Pfaffl et al., 2002b).

Two groups of the ε-estimation can be distinguished:

- Estimation methods based on serial dilution,
- Estimation methods based on a single reaction set-up.

The currently used and partly automated method of determination of ε is the method of serial dilutions (Rasmussen, 2001; Pfaffl and Hageleit, 2001). In this method, serial dilutions of starting template are prepared in those the input nucleic acid concentration is varied over several orders of magnitude. Usually dilution series are prepared by serially diluting the input nucleic acid five to ten times with the sterile water or buffer. Subsequently the CP values are plotted against the natural logarithm of the known start concentration value and ε is estimated as $\varepsilon = 10^{[-1/slope]}$ from the slope of obtained regression line (Rasmussen, 2001). There are some variations of this method, but the serial dilution is always necessary. In the absolute quantification, the

calibration curve can be taken for the ε calculation, provided, it was constructed on the diluted PCR product.

Several ε estimation methods from only a single reaction set-up have been published but not yet integrated into commercial PCR platforms (Wiesner et al., 1992; Liu and Saint, 2002b; Ramakers et al., 2003; Tichopad 2003a). Nonetheless, where raw fluorescence data are available the computation can be performed relatively easily. Active spreadsheet tools based on commercial software such as Excel or Lotus can be helpful here. In praxis it means, that only one sample reaction kinetics is sufficient for the ε determination (Tichopad et al. 2003a). The ε is then determined by fitting the exponentially behaving fluorescence observations with the exponential model [equation 2 or 3]. The erroneous delimitation of the exponentially behaving observations is the main problem here, resulting in a false ε estimation (Tichopad et al. 2003a; Peccoud and Jacob, 1998).

2 MATERIALS AND METHODS

The scope of this chapter is not to provide the reader with a description of myriad of techniques potentially useful for the quantitative real-time RT-PCR, but rather to focus on one method with several moderate modifications. For pragmatic and financial reasons, investigator is often familiar with only one method established and optimised in his laboratory. This method is then possibly slightly modified. Consequently, the sample preparation methods, PCR platform used, fluorescence detection method, or composition and volume of the reaction mix, etc. is often fixed.

The following text provides details of the sample preparation methodology, real-time RT-PCR system, and data acquisition that were used by the author.

2.1 Sample preparation, RNA extraction, and cDNA synthesis

Samples of various ovine and bovine tissues and bovine leucocytes from cell culture were quickly frozen in liquid nitrogen and then stored in –80°C till the total RNA extraction. Subsequently, samples were homogenised and the total RNA was extracted with commercially available preparations peqGOLD TriFast (Peqlab, Erlangen, Germany) or TriPure (Roche, Basel, Switzerland), both utilising a single modified liquid separation procedure (Chomczynski, 1993). RNA pellets were dissolved in water and the concentration was determined spectrophotometrically. Both preparations seemed to give similarly good results (own unpublished observations). No additional purification was performed.

Constant amount of 1000 ng total RNA was reverse–transcribed to cDNA, using 200 units of engineered MMLV H^- Reverse Transcriptase (Promega, Madison, USA), according to the manufacturer's instructions.

Target non-specific priming by random hexamer primers was employed (Zhang and Byrne, 1999).

Also lowered amounts of RNA were occasionally reverse–transcribed without any adjustment of the reaction components. For example, 500 ng of RNA produced corresponding amount of cDNA which, then, in a following PCR quantification did not deviate from other samples derived from 1000 ng (own unpublished observations).

2.2 Real-time RT-PCR on the LightCycler

All PCR runs were performed on the LightCycler instrument (Wittwer et al., 1999; Rasmussen, 2001). Samples belonging to the same group were always run together within one LightCycler run to prevent any inter-run variation. Two approaches of the

RT-PCR were adopted, differing in their separation of the reverse-transcription reaction from the polymerase chain reaction.

2.2.1 Two-step real-time RT-PCR approach

In this approach, the mRNA was separately reverse transcribed into the cDNA on a separate PCR platform. Product of the RT reaction, the cDNA, was then placed into LightCycler capillaries with prepared reaction mix. Details of amplification parameters, primer and amplicon sequence are varying as they had to be enhanced for particular DNA sequences. In general, always 25 ng biological reverse transcribed total RNA or varying experimental concentration of linearised plasmid DNA in 1 µl water were added to 9 µl master mix (i.e. reaction mix without template cDNA). The master-mix was prepared with Fast Start DNA Master SYBR Green I amplifying agent (Roche Diagnostics) according to the manufacturer's instructions. Most of the PCR reactions were carried out in total volume of 10 µl or the reaction was alternatively enhanced to 20 µl. Thirty to forty cycles were applied in various sequences to reach optimal product amount and to generate the full sigmoid fluorescence trajectory. Three or four segment amplification program was constructed including 10 min of initial denaturation at 95°C followed by 3 segment amplification steps; 15 s at 95°C for denaturation, 10 s at respective annealing temperature and 20 s at 72°C for elongation.

Often, a fourth quantification segment was added with sequence-specific temperature above 72°C for quantification purposes. This method, known as the fourth-segment quantification, is often used to ensure higher specificity of product quantified were quantification assay utilises the SYBR Green I. The SYBR Green I is not discriminatory to non-specific DNA product formed during the PCR reaction as long

as it exists as dsDNA. The elevated temperature just below the melting point of the wanted specific product causes other unwanted double stranded DNA in the reaction mix to melt and thus to become 'invisible' (Pfaffl, 2001b).

Eventually, a melting step was performed consisting of 10 s at 95°C, 10s at 60°C and slow heating with a rate of 0.1°C per s up to 99°C with continuous fluorescence measurement. This basic program was occasionally altered for experimental reasons.

2.2.2 One-step real-time RT-PCR approach

In this approach, the RT as well as the PCR reaction was run together on the LightCycler platform using QuantiTect SYBR Green RT-PCR Kit (Qiagen, Hilden, Germany). Prior the PCR temperature program as described above, an RT program of constant 37°C for 20 min was attached. The reaction was set according to the standard protocol recommended by Qiagen, with 5 to 10 ng total RNA.

2.3 Data acquisition and statistical analysis

2.3.1 Quantification data acquisition

The data on amount of amplified sequence in the sample were in form of the pure CP values. CPs were obtained by either the *Fit Point* method (Rasmussen, 2001) or the *Second Derivative Maximum* method (Rasmussen, 2001; Wittwer et al. 1999). In the *Fit Point* method, the threshold level is set subjectively into the exponentially behaving part of the amplification curve. In the *Second Derivative Maximum* method, the positive maximum of the second derivative of the amplification curve is computed as the threshold level. Fractional number of cycles at this threshold level is the CP value (figure 1).

2.3.2 Fluorescence data acquisition

The fluorescence data was taken directly from the LightCycler software (various versions). The data is produced by repeated measurements of the fluorescence emitted by the reaction system in the capillary, utilising SYBR Green I – a double stranded DNA (dsDNA) intercalating binding dye. The fluorescence acquisition was done either at the end of the elongation segment or at the end of the appended fourth segment.

2.3.3 Statistical tests

Once any sort of the above mentioned data had been acquired, a statistical test was applied to confirm or reject the study assertion (e.g. hypothesis). A statistical test is a procedure for deciding whether hypothesis about a quantitative feature of a *general* population is true or false (e.g expression of prion-protein gene in the neocortex is *generally* higher then in the muscle) . We test an hypothesis of this sort by drawing a random sample from the population in question and calculating an appropriate statistic on its items. That is, only some randomly chosen samples are to report about the entire population.

A statistical test is based on a probability level alpha (α). It indicates the probability of rejecting the statistical hypothesis tested when in fact, that hypothesis is true. Before conducting any statistical test, it is important to establish a value for α . For most biological, and for many other scientific purpose, it is customary to set α at 0.05.

It is always necessary to decide what statistics to use, what sample size to employ and what criteria to establish for rejection of the hypothesis tested. One- and two-way ANOVA models were satisfactory tools to analyse PCR data. Where multiple pair-

wise comparisons between groups were done, the Tukey method of the overall α value adjustment (Steel and Torrie, 1980) was adopted. Provided, the data had the Gaussian distribution, no transformation was taken. Alternatively, the data was *log* transformed.

3 RESULTS AND DISCUSSION

After optimisation, all real-time PCR assays could be routinely run generating specific amplicons, showing no primer dimers, single sharp peak, identical melting points (Ririe, 1997) and expected lengths on the agarose gel.

The PCR is a complex method of rather cumbersome exponential than a straightforward linear character. This bears an important inherent disadvantage in it, because also any error is amplified in the exponential fashion together with the product (Peccaud and Jacob, 1996). This can result in a great under-/over-estimation of measured concentration of analyt. Many avoidable sources of error are already present in the initial reaction mix (Rossen et al., 1992; Wilson, 1997; Tichopad et al., 2004) and/or in the surrounding conditions (e.g all chemical or mechanical reaction inhibitors, integrity of RNA, loading error or differences in temperature in the lab). If two samples of RNA come from different tissues, the present tissue-specific contaminants cause discrepancy at the output of the real-time RT PCR (Tichopad et al.; 2004).

As long as the quantification takes place within the exponentially behaving phase of the second portion, problems associated with its true sensitivity, reproducibility and specificity are minimised (Tichopad et al., 2003a). The quantification event should be conducted possibly soon, within the early exponential phase.

Theoretically, comparable samples should produce comparable amplification curves. Watching the amplification course on the monitor provides the first hint to decision on the assay's performance. Unfortunately, any judgement on the amplification curve is heavily arbitrary as no parameters of the amplification curve are produced directly by the platform's software up to now. To get a better insight into the trajectory of the amplification, some mathematical models such as in Tichopad et al. (2002 and 2004) suggested four-parametric sigmoid model and the four parametric logistic model (Tichopad et al. 2003a) can be useful. These full-trajectory models give a smooth amplification curve. Unlike in Tichopad et al. (2002) where the modelled amplification trajectory was used for *optimisation* of the reaction conditions, it could also be used for *validation* of the quantification assay's performance consisting of several samples being compared (Tichopad et al., 2004). Various curves for different samples are obtained in this way. Parameters of these curves can be statistically compared, and the first information about sample's comparability can be extracted in this way. Similarly, the parameters from the smoothing model can be used to analyse an effect of any substance on the polymerase performance. In Tichopad et al.[2] the inhibition of the polymerase enzyme by tea polyphenols was shown.

The four-parametric sigmoid model's parameter ***b*** detects any dissimilarity between samples. This detection is more sensitive than the ε computation and comparison between samples. This is because the model fits the full data, whereas ε is computed from five to ten fluorescence observations only. From the parameters of the smoothing full-data models [equations 4 and 9] the parameters ***a*** and ***b*** are of the greatest importance because they report about the inhibition of the reaction. Parameter

[2] **Tichopad, A.**, Polster, J, Pecen, L. & Pfaffl, W. Inhibition of Taq Polymerase and MMLV Reverse Transcriptase performance in presence of tea polyphenols (+)-Catechin and (-)-Epigallocatechin-3- Gallate (EGCG). Journal of Ethnopharmacology, (Submitted). *Attached in appendix*

n_0, as well as the CP, reports about the starting concentration of DNA in sample. Both ***a*** and ***b*** model parameters can be obtained sample-specific, so that no additional PCR runs with serial dilutions must be done. Where incomparable samples are assumed, a standardisation procedure based on knowledge of amplification efficiencies is to be adopted (Pfaffl, 2001a; Meijerink, 2001; Pfaffl et al., 2002b; Liu and Saint, 2002a). Some standardisation procedures can be adopted from micro-array technologies, where often similar problems are faced (Schuchhardt et al. 2000).

To correct for discrepancy between ideal and real conditions in reaction, the reaction's reported amplification efficiency ε must be estimated (Pfaffl, 2001a; Liu and Saint, 2002a; Ramakers et al., 2003). Since the amplification trajectory is known to behave exponentially in its first phase of the portion 2 (figure 2), the equation [3] can be employed as a smoothing model to extract the parameter ε. Suggested method of estimation from a single reaction set-up as presented by Tichopad et al. (2003a) is fully instrumental, with no decision step necessary to be done by investigator. For this reason, there is no subjective bias introduced into the CP estimator. This method returns a sample-specific amplification efficiency estimation value. This value reflects only the exponentially behaving part of the amplification curve and it could be the second parameter beside the crossing point generated by a real-time PCR platform. Knowledge of the sample-specific ***CP*** and ε value would increase the accuracy of the real-time PCR assay in both quantification models [equation 10 and 11]. Method of kinetic outlier detection (KOD) described by Bar et al. (2003) is an ideal tool to find samples those impair the assay accuracy due to dissimilar ε.

A method of standardisation of gene-expression by grouped index constructed as the geometric mean of CP values of several housekeeping genes was presented by Vandesompele et al. (2002) and Pfaffl et al. (2004). Similarly to Pfaffl et al. (2004),

cluster analysis based on the correlation matrix was suggested by Tichopad et al.[3] In contrast to standardisation by a single gene, these methods have a great advantage of standardisation by a robust basis. Employing such an approach, the above mentioned problems with a possible standard gene regulation can be avoided. The computing procedure by Pfaffl et al. (2004) can also integrate other genes then just housekeeping genes into the groped index. If these genes show stable expression comparable to housekeeping genes, they can be used as a standard. The *BestKeeper* software can compare expression levels of up to 10 housekeeping genes together with 10 target genes, each of up to 100 samples. It determines the 'optimal' housekeeping genes and calculates the geometric mean of the 'best' suited ones, employing the pair-wise correlation analysis of all pairs of candidate genes. The earlier presented *GeNorm* software (Vandesompele et al. 2002) is restricted to the housekeeping genes analysis only, whereas, in the *BestKeeper* software up to 10 target genes can also be analysed.

Alternatively to the standardisation by another gene, an example of absolute quantification with calibration curve is shown in figure 1 (Tichopad et al. 2003b). Several such samples of known concentration diluted with a constant dilution step produce a calibration curve. The initial concentration of unknown biological sample can be then easily obtained from the calibration curve if the CP is known. Neocortex as a tissue from central nerve system is known to be affected by the pathogen form of prion protein (Prusiner, 1998). Also here, its amplification curve rises up and generates its CP sooner then the curve of the liver.

Finally, a complex result of a real-time PCR platform should include not only the CP of a given sample, but also parameters of its amplification curve as given by

[3] **Tichopad, A.**, Pfaffl, M.W. & Pecen, L. Distribution-insensitive rank-order dissimilarity measure based clustering on real-time PCR data of potential gene expression normalization candidate genes. Journal of Bioinformatics and Computational Biology, (Submitted). *Attached in appendix*

smoothing model and the sample specific amplification efficiency. This would surely offer more robust data fundament for an analysis of the gene-expression.

The t-test or ANOVA can by employed to estimated difference between treatment groups. For non-normally distributed data some non-parametric tests such as Wilcoxon rank-sum test for two-sample data or Kruskal Wallis test for multisample data should be used. Another possibility to deal with non-normally distributed data is to employ parametric tests on transformed data. Transformation such as logarithm or sinus of the raw data can improve distribution. Statistical models dealing with random effects should be applied where more PCR runs or repeated RT procedure is needed to complete the experiment. So called mixed models can model the covariance structure caused by repeated experimental design (Littell et al., 1998).

In case of tissue expression pattern studies, variability between the assay's performances in individual samples derived from various tissues is to be assumed (Pfaffl et al., 2001; Tichopad et al. 2004). Further, also RT reaction performance and selective post-mortal changes are to be expected in different tissue samples. The quantification with calibration curve will yield some unavoidable error due to heterogeneous performance between tissue-derived samples and samples and calibration curve. The different-tissue-derived samples can hardly be standardised and compared as the reference gene can vary in its expression between the tissues. For above-mentioned reasons, expression-pattern studies are problematic and their results should be considered with caution.

Physiological changes in organism can be quantified relatively in samples derived from the same tissue type. The tissue-specific disturbance is not relevant here. The *pre-treatment vs. post-treatment* experiments are performed on the same material. Nevertheless, some discrepancy can be also present in samples from the same

biological material. Reasons can be various; residua of the treatment agent, varying sampling procedure, extraction times, varying temperatures during sampling or extraction etc.. Very sensitive are samples obtained from cultured cells where varying proportion of dead cell in medium, changes in composition of medium, and differing sampling volumes are responsible for heterogeneous assay (own unpublished data).

The CPs of samples with the homogeneous assay performance could be compared without any additional standardisation (equation 5 and 10).

For overall improvement of the entire evaluation process the flow chart of real-time RT-PCR data evaluation is suggested here (figure 5) with following steps: The fluorescence data are plotted against the cycle number *n*. Then a smoothing model is fitted, producing amplification curves. The CP values are obtained in a non-arbitrary computational way either directly from the PCR platform or they are computed from the smoothed model as its second or generally n^{th} derivative maximum. Amplification efficiency $\boldsymbol{E}$ is estimated for each individual sample by fitting exponential model [equation 3] into exponentially behaving phase of the amplification course. Parameters of amplification curves are then compared, using some two-sample or multiple-sample statistical test, testing whether they are homogeneous or heterogeneous between compared groups. Where there are no differences in PCR performance between compared groups the relative quantification can be performed without employing any internal standard [equation 10]. Alternatively, if there are differences in PCR performance between compared groups, the relative quantification must be standardized by stable expressed internal standard [equation 11].

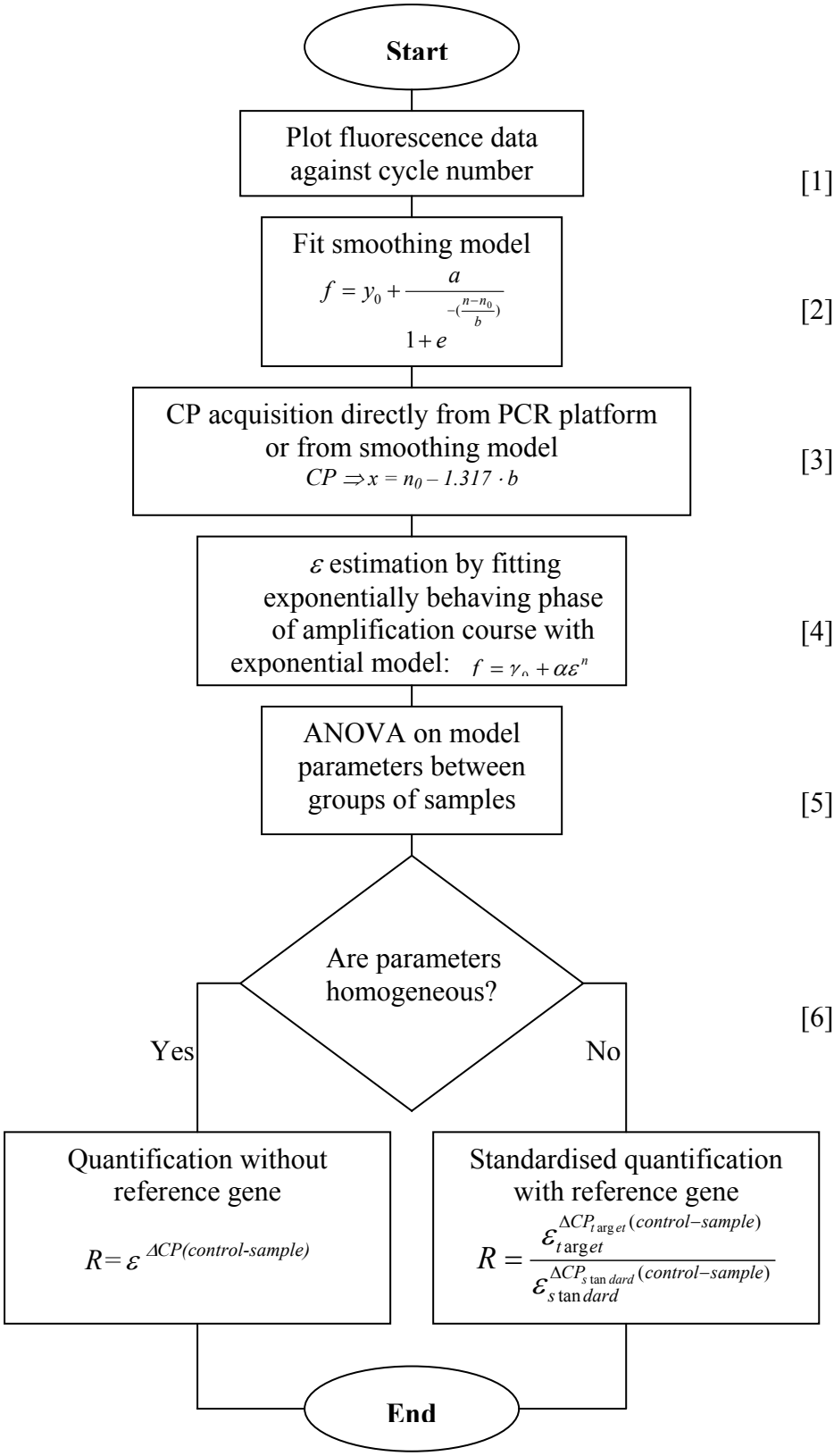

Start
Plot fluorescence data against cycle number
[1]
Fit smoothing model
$f = y_0 + \frac{a}{1+e^{-(\frac{n-n_0}{b})}}$
[2]
CP acquisition directly from PCR platform or from smoothing model
$CP \Rightarrow x = n_0 - 1.317 \cdot b$
[3]
ε estimation by fitting exponentially behaving phase of amplification course with exponential model: $f = \gamma_n + \alpha\varepsilon^n$
[4]
ANOVA on model parameters between groups of samples
[5]
Are parameters homogeneous?
[6]
Yes
No
Quantification without reference gene
[7]
$R = \varepsilon^{\Delta CP(control-sample)}$
Standardised quantification with reference gene
$R = \frac{\varepsilon_{target}^{\Delta CP_{target}(control-sample)}}{\varepsilon_{standard}^{\Delta CP_{standard}(control-sample)}}$
[8]
End

Figure 5. Flow chart of real-time RT-PCR relative data evaluation.

[1] Fluorescence data are plotted against the cycle number ***n***

[2] Fluorescence data are fitted with smoothing model – amplification curves are obtained. Here as an example the four-parametric sigmoid model is used [equation 4].

[3] CP values are obtained in a non-arbitrary computational way either directly from the PCR platform or they are computed from the smoothed model as its second or n^{th} derivative maximum.

[4] ε is estimated for each individual sample by fitting exponential model [equation 3] into exponentially behaving phase of the amplification course.

[5] Parameters of amplification curves are compared by ANOVA test whether they are homogeneous or heterogeneous between compared groups of samples.

[6] Decision on the predetermined probability level, whether the parameters of the amplification curves between compared groups of samples are homogeneous or not.

[7] There are no differences in PCR performance between compared groups of samples as shown by ANOVA test. Relative quantification can be performed without employment of internal standard [equation 10].

[8] There are differences in PCR performance between compared groups of samples as shown by ANOVA test. Relative quantification must be standardized by stable expressed internal standard [equation 11].

4 CONCLUSION

The novel monitoring of a fluorescence emitted by the dye-product intercalation during real-time polymearase chain reaction produces a number array that is an important source of additional quantitative information. Amplification trajectories of individual PCR samples can be visualised out of this observations and subsequently analysed. Any heterogeneity in sample performance other then due to different starting template concentrations introduces error into results. The amplification trajectory is a non-linear, rather logistic than exponential, posing a conflict to the recent quantification methods based on the assumed exponential character of PCR. The real exponential trend must be therefore detected and quantified.

To address the above problems mathematical models were suggested for describing the full amplification trajectory and disclosing heterogeneity between samples. Statistical diagnostic procedure was suggested for stepwise fitting background fluorescence observations with the linear regression model with subsequent residual diagnostics. This procedure can reliably inspect the reaction's exponential trend. It was further suggested here that the quantification step be carried out as early as possible to take advantage of the exponential fashion of amplification. In praxis it means, that the first observation detectable on the trajectory just above the background phase gives the best threshold level for quantification decisions. This helps to minimise error caused by reaction's deviation from the exponential. Where heterogeneity in reaction performance is present, a good standardisation method must be applied. It was shown that computing correlation matrix for all assayed candidate genes could point out suitable standards also including some non-regulated target genes. Standardisation index can be computed as geometric mean of the successful candidates.

ACKNOWLEDGEMENTS

The Author thanks his supervisor PD Dr. M. W. Pfaffl and Prof. Dr. H.H.D. Meyer for their initiating this thesis, supervision, and critical reviews.

Furthermore, author thanks all internal and external colleagues those made his work on this thesis possible; particularly Dr. A Didier, Dipl. Ing. agr. C. Prgomet, Dipl. Ing. agr. M. Dilger, MSc. A. Dzidic, Dr. T. Neuvians, Doc. Dr. L. Pecen, Dr. G. Schwarz, D. Tetzlaff and Prof. Dr. J. Polster.

REFERENCES

Bar, T., Stahlberg, A., Muszta, A., Kubista, M. (2003) Kinetic Outlier Detection (KOD) in real-time PCR. Nucleic Acids Res 31, e105.

Bustin, S.A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 25, 169-193.

Chang, T.J., Juan, C.C., Yin, P.H., Chi, C.W., Tsay, H.J. (1998) Up-regulation of beta-actin, cyclophilin and GAPD in N1S1 rat hepatoma. Oncol Rep 5, 469-471.

Chomczynski, P.A. (1993). Reagent for the single–step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. Biotechniques 15, 532–534.

Cunnick, G., Jiang W.G. Quantitation of lymphangiogenesis using the iCycler iQ real-time PCR detection system and Scorpions detection system, Bio-Rad bulletin 2806. Available under: http://www.bio-rad.com/LifeScience/pdf/Bulletin_2806.pdf

de Silva, D., Wittwer, C.T. (2000) Monitoring hybridization during polymerase chain reaction. J Chromatogr B Biomed Sci Appl 741, 3–13.

de Stefano, J.J.; Buiser, R.G., Mallaber, L.M., Myers, T.W., Bambara, R.A., Fay, P.J. (1991) Polymerization andRNase H activities of the reverse transcriptase from avian myeloblastosis, human immunodeficiency, and Moloney murine leukemia viruses are functionally uncoupled. J Biol Chem 266, 7423-7431.

Ferré, F. (1992). Quantitative or semi–quantitative PCR: Reality versus Myth. PCR Methods and Applications 2, 1–9.

Foss, D.L., Baarsch, M.J., Murtaugh, M.P. (1998) Regulation of hypoxanthine phosphoribosyltransferase, glyceraldehyde-3-phosphate dehydrogenase and beta-actin mRNA expression in porcine immune cells and tissues. Anim Biotechnol. 9, 67-78.

Freeman, W.M., Walker, S.J., Vrana, K.E. (1999) Quantitative RT–PCR: pitfalls and potential. BioTechniques 26, 112–125.

Gibson, U.E., Heid, C.A., Williams, P.M. (1996). A novel method for real time quantitative RT–PCR. Genome Res 6, 995–1001.

Higuchi, R., Fockler, C., Dollinger, G., Watson, R. (1993) Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions. Biotechnology 11, 1026–1030.

Hocquette, J-F., Brandstetter, A.M. (2002) Common practice in molecular biology may introduce statistical bias and misleading biological interpretation. J Nutr Biochem 13, 370-377.

Holland, P.M., Abramson, R.D., Watson, R., Gelfand, D.H. (1991) Detection of specific polymerase chain reaction product by utilizing the 5' – 3' exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci USA 88, 7276–7280.

Kainz, P. (2000) The PCR plateau phase – towards an understanding of its limitations. Biochim Biophys Acta 1494, 23–27.

Klein, D. (2002) Quantification using real-time PCR technology: Applications and limitations. Trends Mol Med 8, 257–260.

Larrick, J.W. (1992) Message amplification phenotyping. Trends Biotechnol 10, 146-152.

Liss, B. (2002) Improved quantitative real-time RT–PCR for expression profiling of individual cells. Nucleic Acids Res 30, e89.

Littell, R.C., Henry, P.R., Ammerman, C.B. (1998) Statistical Analysis of Repeated Measures Data Using SAS Procedrures. J Anim Sci 76, 1216-1231.

Liu, W., Saint, D.A. (2002a) A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. Anal Biochem 302, 52–59.

Liu, W., Saint, D.A. (2002b) Validation of a quantitative method for real time PCR kinetics. Biochem Biophys Res Commun 294, 347+353.

Livak, K.J., (2001) ABI Prism 7700 Sequence detection System User Bulletin #2 Relative quantification of gene expression:

Livak, K.J., Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(–Delta Delta C(T)). Methods 25, 402–408.

Mannhalter, C., Koizar, D., Mitterbauer, G. (2000) Evaluation of RNA isolation methods and reference genes for RT-PCR analyses of rare target RNA. Clin Chem Lab Med 38, 171-177.

Marras, S.A.E., Kramer, F.R., Tyagi, S. (1999) Multiplex detection of single–nucleotide variations using molecular beacons. Genet Anal 14, 151–156.

Meijerink, J., Mandigers, C., van de Locht, L., Tonnissen, E., Goodsaid, F., Raemaekers, J. (2001) A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR. J Mol Diagn 3, 55–61.

Meuer, S., Wittwer, C. Nakagawara, K. (2001) Rapid cycle real – time PCR: Methods and Applications (Springer, Heidelberg, 2001).

Morrison, T.B., Weis, J.J., Wittwer, C.T. (1998). Quantification of low–copy transcripts by continuous SYBR Green I monitoring during amplification. Biotechniques 24, 954–8.

Muller, P.Y., Janovjak, H., Miserez, R., Dobbie, Z. (2002) Processing of Gene Expression Data Generated by Quantitative Real-Time RT-PCR. BioTechniques 32, 2-7.

Orlando, C., Pinzani, P., Pazzagli, M. (1998) Developments in quantitative PCR. Clin Chem Lab Med 36, 255–269.

Peccoud, J., Jacob, C. (1996) Theoretical uncertainty of measurements using polymerase chain reaction. Biophys J 71, 101-108.

Peccoud, J., Jacob, C. (1998) Statistical estimation of PCR amplification rates. In: Gene Quantification (ed. Ferré, F.) 111–128 (Birkhauser, Boston).

Pfaffl, M.W., **Tichopad, A.**, Prgomet, Ch. & Neuvians, T. (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. Biotechnol Lett 26, 509-515.

Pfaffl, M.W., Hageleit, M. (2001) Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in Real-time RT-PCR. Biotechnol Lett 23, 275-282.

Pfaffl, M.W. (2001a) A new mathematical model for relative quantification in Real-time RT-PCR. Nucleic Acids Res 29, e45.

Pfaffl, M.W. (2001b) Development and validation of externally standardised quantitative Insulin like growth factor–1 (IGF–1) RT–PCR using LightCycler SYBR ® Green I technology. In: Rapid cycle real – time PCR: Methods and Applications (eds. Meuer, S., Wittwer, C., Nakagawara, K.) 21–34 (Springer, Heidelberg).

Pfaffl, M.W., Horgan, G.W.; Dempfle, L. (2002b) Relative Expression Software Tool (REST©) for group wise comparison and statistical analysis of relative expression results in Real-time PCR. Nucleic Acids Res 30, e36.

Pfaffl, M.W., Lange, I.G., Daxenberger, A., Meyer, H.H. (2001) Tissue–specific expression pattern of estrogen receptors (ER): quantification of ER alpha and ER beta mRNA with real-time RT–PCR. Acta Pathologica Microbiologica et Immunologica Scandinavica 109, 345–55.

Prusiner, S.B. (1998) Prions. Proc Nat Acad Sci USA 95, 13363-13383.

Ramakers, C., Ruijter, J.M., Lekanne Deprez, R.H.L., Moorman, A.F.M. (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 339, 62-66.

Rasmussen, R. (2001). Quantification on the LightCycler instrument. In Rapid cycle real-time PCR: Methods and Applications (Meuer, S., Wittwer, C. & Nakagawara, K., eds) Pp. 21–34. Heidelberg: Springer.

Ririe, K.M., Rasmussen, R. and Wittwer, C.T. (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Anal Biochem 245, 154–160.

Rossen, L., Norskov, P., Holmstrom, K., Rasmussen, F.O. (1992). Inhibition of PCR by components of food sample, microbial diagnostic assay and DNA–extraction solutions. Int J Food Microbiol 17, 37–45.

Serazin-Leroy, V., Denis-Henriot, D., Morot, M., de Mazancourt, P., Giudicelli, Y. (1998) Semi-quantitative RT-PCR for comparison of mRNAs in cells with different amounts of housekeeping gene transcripts. Mol Cell Probes 12, 283-291.

Schmittgen, T.D., Zakrajsek, B.A. (2000) Effect of experimental treatment on housekeeping gene expression: validation by real - time, quantitative RT-PCR. J Biochem Biophys Methods 46, 69-81.

Schmittgen, T.D. (2001) Real-time quantitative PCR. Methods 25, 383–385.

Schmittgen, T.D., Zakrajsek, B.A., Mills, A.G., Gorn, V., Singer, M.J., Reed, M.W. (2000) Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. Anal Biochem 285, 194-204.

Schnell, S., Mendoza, C. (1997) Theoretical description of the polymerase chain reaction. J Theor Biol 188, 313-318.

Schuchhardt, J., Beule, D., Malik, A., Wolski, E., Eickhoff, H., Lehrach, H. Herzel, H. (2000) Normalisation strategies for cDNA microarrays. Nucleic Acids Res 28, e47.

Souazé, F., Ntodou-Thomé, A., Tran, C.Y., Rostene, W., Forgez, P. (1996) Quantitative RT-PCR: Limits and accuracy. BioTechniques 21, 280-285.

Steel, R.G.D., Torrie, J.H. (1980) Principles and procedures of statistics - a biometrical approach. Second edition. McGraw-Hill Book Company, New York, USA. 633 pp.

Starnbach, M.N., Falkow, S., Tomkins, S.L. (1989). Species-specific detection of Legionella pneumophila in water by DNA amplification and hybridization. J Clin Microbiol 27, 1257–1261.

Suzuki, T., Higgins, P.J., Crawford, D.R. (2000) Control Selection for RNA Quantitation. BioTechniques 29, 332-337.

Swift, G.H., Peyton, M.J., macDonald, R.J. (2000) Assessment of RMA quality by semi-quantitative RT-PCR of multiple regions of a long ubiquitous mRNA. Biotechniques 28, 524-531.

Swabe, H., Stein, , U., Walther, W. (2000) High-copy cDNA amplification of minimal total RNA quantities for gene expression analysis. Mol Biotechnol 14, 165-172.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., Heinen, E. (1999) Housekeeping genes as internal standards: use and limits. J Biotechnol 75, 291-295.

Tichopad, A., Didier, A., Pfaffl, M.W. (2004) Inhibition of real-time RT-PCR quantification due to tissue-specific contaminants, Molecular and Cellular Probes 18, 45-50.

Tichopad, A., Dilger, M., Schwarz, G., Pfaffl, M.W. (2003a) Standardized determination of real-time PCR efficiency from a single reaction set-up. Nucleic Acids Res 31, E122.

Tichopad, A., Dzidic, A., Pfaffl, M.W. (2002) Improving quantitative real-time RT-PCR reproducibility by boosting primer-linked amplification efficiency. Biotechnol Lett 24, 2053-2056.

Tichopad, A., Pfaffl, M.W., Didier, A. (2003b) Tissue-specific expression pattern of bovine prion gene: quantification using real-time RT-PCR. Mol Cell Probes 17, 5-10.

Vandesompele, J., de Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F. (2002) Accurate normalisation of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Gen Biol 3, 1-12.

Warrington, J.A., Nair, A., Mahadevappa, M., Tsygantskaya, M. (2000) Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiol Genomics 2, 143-147.

Whitcombe, D., Theaker J., Guy, S.P., Brown, T., Little, S. (1999) Detection of PCR products using self–probing amplicons and flourescence. Nature 17, 804–807.

Wiesner, R.J., Ruegg, J.C., Morano, I. (1992) Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissue. Biochem Biophys Res Commun 183, 553-559.

Wilson, I.G. (1997) Inhibition and Facilitation of Nucleic Acid Amplification. Appl Environ Microbiol 63, 3741–3751.

Wilson, I.G., Gilmour, A., Cooper, J.E. (1993) Detection of toxigenic microorganisms in foods by PCR. In: New techniques in food and beverage microbiology (eds. Kroll, R.G., Gilmour, A., Sussman, M.) 163–172 (Blackwell Publishers, London).

Wittwer, C.T., Ririe, K. M., Andrew, R.V., David, D.A., Gundry, R.A., Balis, U.J. (1997) The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. Biotechniques 22, 176-181.

Wittwer, C.T., Gutekunst, M., Lohmann, S. (1999) Method for quantification of an analyte. United States Patent No.: US 6,303,305 B1.

Wong, L., Pearson, H., Fletcher, A., Marquis, C.P., Mahler, S. (1998) Comparison of the Efficiency of Moloney Murine Leukemia Virus (M-mulv) Reverse Transcriptase, rnase H—M-mulv Reverse Transcriptase and Avian Myeloblastoma Leukemia Virus (AMV) Reverse Transcriptase for the Amplification of Human Immunoglobulin Genes. Biotechnology Techniques 12, 485-489.

Zhang, J., Byrne, C.D. (1999) Diferential priming of RNA templates during cDNA synthesis markedly affects both accuracy and reproducibility of quantitative competitive reverse-transcriptase PCR. Biochem J 337, 231-241.

ALEŠ TICHOPÁD – LEBENSLAUF

Silberhornstraße. 3
D-85551 Kirchheim b. München

Email: alestichopad@yahoo.de

Geburtsdatum: 20.05.1975
Geburtsort: Ostrau, Tschechische Republik
Staatsangehörigkeit: Tschechisch

Schulbildung:

1981-1989 Grundschule

1989-1993 Höhere Technische Lehranstalt

Mai 1993 Abitur

1993-1994 Kurse: Biologie, Chemie, Mathematik und Englisch an der Karls-Universität in Prag und der Technischen Universität in Ostrau

Ausbildung:

1995-2001 Karls-Universität in Prag, Fakultät für Naturwissenschaften, Studienfach Biologie, Spezialisierung für Ethologie und Ökologie

Thema der Diplomarbeit: *Mehrjährige Wachstumstrends bei der Getreidelaus- Population*

Januar 2001 Studium erfolgreich abgeschlossen mit dem Titel *Magistri Biologia* (entspricht *Dipl. Biol.*)

März - April 2001 Praktikum am Institut für Physiologie der TUM: Molekularbiologische Methoden

April 2001 Beginn der Promotion am Institut für Physiologie der TUM

Thema der Dissertation: *Real-time RT-PCR transcriptomics: Improvement of evaluation methods*

Forschungstätigkeiten:

- Entwicklung und Anwendung von Methoden für Auswertungen von Genexpressionsdaten
- Quantifizierung der Genexpression des bovinen Prions und das BSE Übertragungsrisiko in verschiedenen Geweben des Rindes
- Untersuchung der Einflüssen von Tee Polyphenolen auf die Polymerase und Reverse-Transkriptase

- Etablierung eines transfizierten bovinen Zellkulturmodels für die Untersuchung des Zytokine-Crosstalks zwischen Epithelzellen und Leukozyten

Beruf: **Seit März 2003** angestellt als Biometriker in der internationalen Auftragsforschungsorganisation IMFORM GmbH

Sprachenkenntnisse: Deutsch (Verhandlungssicher)
Englisch (Verhandlungssicher)
Tschechisch (Muttersprache)

Aleš Tichopád 25. Oct. 2002

ALEŠ TICHOPÁD –LIST OF PUBLICATIONS

Full length papers

Pfaffl, M.W., Tichopad, A., Prgomet, Ch. & Neuvians, T. (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel spreadsheet tool using pair-wise correlations, *Biotechnology Letters* 26, 509-515.

Tichopad, A., Didier, A. & Pfaffl, M.W. (2004) Inhibition of real-time RT-PCR quantification due to tissue-specific contaminants, *Molecular and Cellular Probes* 18, 45-50.

Tichopad, A., Dilger, M., Schwarz, G. & Pfaffl, M.W. (2003) Standardized determination of real-time PCR efficiency from a single reaction set-up, *Nucleic Acids Research* 31, e122.

Tichopad, A., Pfaffl, M.W. & Didier, A. (2003) Tissue-Specific Expression Pattern of Bovine Prion: Quantification Using Real-Time RT-PCR, *Molecular and Cellular Probes* 17, 5-10.

Tichopad, A., Dzidic, A. & Pfaffl, M.W. (2002) Improving quantitative real-time RT-PCR reproducibility by boosting primer-linked amplification efficiency, *Biotechnology Letters* 24, 2053-2056.

Manuscripts submitted for publication

Tichopad, A., Polster, J, Pecen, L. & Pfaffl, W. Inhibition of Taq Polymerase and MMLV Reverse Transcriptase performance in presence of tea polyphenols (+)-Catechin and (-)-Epigallocatechin-3- Gallate (EGCG). *Journal of Ethnopharmacology*, (Submitted).

Tichopad, A., Pfaffl, M.W. & Pecen, L. Distribution-insensitive rank-order dissimilarity measure based clustering on real-time PCR data of potential gene expression normalization candidate genes. *Journal of Bioinformatics and Computational Biology*, (Submitted).

Posters

Tichopad, A. & Pfaffl, W. (2004) Search by cluster analysis for steadily expressed genes with application as normalization index in real-time RT-PCR (1st International qPCR Symposium & Application Workshop © Transcriptomics, Clinical Diagnostics & Gene Quantification, 3rd - 6th March, 2004 in Freising-Weihenstephan, Germany).

Tichopad, A. Polster, J. & Pfaffl, W. (2003) Inhibition of Taq Polymerase and MMLV Reverse Transcriptase performance in presence of polyphenolic compounds: (+)-Catechine & Epigallocatechin Gallate (EGCG) (1st International Conference on Polyphenols and Health, 18th – 21th November, 2003 in Vichy, France).

Full Length Papers

Biotechnology Letters **26:** 509–515, 2004.

Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations

Michael W. Pfaffl*, Ales Tichopad, Christian Prgomet & Tanja P. Neuvians
Physiology, FML-Weihenstephan, Centre of Life and Food Science, Technical University of Munich, Germany
**Author for correspondence (Fax: +49 8161 714204; E-mail: pfaffl@wzw.tum.de)*

Received 3 December 2003; Revisions requested 23 December 2003; Revisions received 20 January 2004; Accepted 22 January 2004

Key words: β-actin, housekeeping gene, PCR normalisation, RT-PCR, somatotropic axis, ubiquitin

Abstract

The stability of standard gene expression is an elementary prerequisite for internal standardisation of target gene expression data and many so called housekeeping genes with assumed stable expression can exhibit either up- or down-regulation under some experimental conditions. The developed, and herein presented, software called *BestKeeper* determines the best suited standards, out of ten candidates, and combines them into an index. The index can be compared with further ten target genes to decide, whether they are differentially expressed under an applied treatment. All data processing is based on crossing points. The *BestKeeper* software tool was validated on four housekeeping genes and 10 members of the somatotropic axis differentially expressed in bovine *corpora lutea* total RNA. The *BestKeeper* application and necessary information about data processing and handling can be downloaded on http://www.wzw.tum.de/gene-quantification/bestkeeper.html

Introduction

Reporting of the amount of target mRNA requires an accurate template preparation and relevant standardisation (Pfaffl 2001). This affects more advanced methods of gene expression study such as real-time PCR (Pfaffl 2001) or microarrays (Schuchhardt *et al.* 2000), as well as the traditional blotting methods. Since several parameters of the quantification procedure (e.g. inhibitory factors of the tissue, integrity of the RNA, loading error, enzyme or primer performance, etc.) must be controlled, numerous standardisation methods have been proposed (Suzuki *et al.* 2000, Thellin *et al.* 1999, Vandesompele *et al.* 2002). In most of them, just a distinct part of the whole real-time RT-PCR quantification procedure is reflected. For example, if the raw expression data is standardised to the amount of biological material, then the inhibitory, tissue-born residua present in sample will be disregarded. Similarly, if the quantification data is expressed per amount of total RNA extracted, then the predominant ribosomal RNAs (5S, 18S and 28S), known to vary in their proportion in the total RNA, can cause significant shifts in the results. This means that a 'full procedure control' is necessary.

In the relative quantification (Serazin-Leroy *et al.* 1998), the standardisation with another gene, whose expression is believed to be constant, is the method of choice (Suzuki *et al.* 2000, Thellin *et al.* 1999). The sequence of the standard and the target template are present in the sample during the whole assay. Therefore, the standard mimics all disturbances of the target sequence. A myriad of housekeeping genes (HKG), such as tubulins, actins, glyceraldehyde-3-phosphate dehydrogenase (GAPD), albumins, cyclophilin, micro-globulins, ribosomal units (18S or 28S rRNA), ubiquitin (UBQ) have been described. On the other hand, some of these genes has been reported to be regulated occasionally (Foss *et al.* 1998, Schmittgen & Zakrajsek 2000). Taking the above-mentioned arguments into account, one must con-

clude that there is no absolutely ideal way to control disturbances in the quantification procedure.

Before any gene is chosen as a standard, an exhaustive search is needed to ensure that no significant regulation occurs. This can, however, be a circular problem, as the expression data of the tested standard, as well, has to be standardised. A possible solution might be a use of more than just one HKG in a form of weighted expression index. To address this problem, an Excel based spreadsheet software application named *BestKeeper* was established and tested on biological material.

Materials and methods

Collection of bovine Corpora lutea

Thirty-one cows at the mid-luteal phase (days 8–12) were injected intra muscularly with 500 μg prostaglandin (PG) F2α analogue, *Cloprostenol* (Estrumate, Intervet, Germany). *Corpora lutea* (4–5 per group) were collected by trans-vaginal ovariectomy at six intervals after PGF2α-injection. Five control *corpora lutea* were randomly collected from untreated cows at the mid-luteal phase. All *corpora lutea* were aliquoted, immediately frozen in liquid N_2 and than stored at $-80\,^{\circ}$C until RNA extraction.

Total RNA extraction

The total RNA was extracted from 100 mg slices of deep frozen tissue with the peqGOLD TriFast™ (PeqLab, Erlangen, Germany), utilising the single step modified liquid separation procedure (Chomczynski 1993). The integrity of the total RNA was determined by electrophoresis on 2% (w/v) agarose gels. Nucleic acid concentrations were measured at 260 nm. Purity of the total RNA extracted was determined as the 260 nm/280 nm ratio with expected values between 1.8 and 2.

Two step RT real-time PCR

One μg total RNA was reverse-transcribed to cDNA in 40 μl volume in the Mastercycler Gradient (Eppendorf, Hamburg, Germany) thermal cycler. Following reaction mix was set: RT buffer (50 mM Tris, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 10 mM DTT and 300 μM dNTPs. The RNA was first denaturated at 65 °C for 5 min. For the subsequent RT reaction, 100 μM random hexamer primers (MBI Fermentas, St. Leon-Rot, Germany), 200 units M-MLV H^-, Reverse Transcriptase (Promega, Madison, USA), and 12.5 U RNase inhibitor (Roche Diagnostics, Mannheim, Germany) were added and the reaction incubated at 42 °C for 60 min. Eventually, samples were heated for 1 min at 99 °C to terminate the RT reaction.

Primer sequences of UBQ, GAPD, β-actin, 18S rRNA, IGF-1 (insulin-like growth factors type 1), IGF-2, IGFR-1 (insulin-like growth factor receptor type 1), IGFR-2, IGFBP-1 (insulin-like growth factor binding protein type 1) – IGFBP-6 were designed to span at least one intron (Pfaffl *et al.* 2002). Primers were synthesized commercially (MWG Biotech, Ebersberg, Germany). PCR conditions were optimised on the gradient thermal cycler and on the LightCycler (Roche Diagnostic). Real-time PCR using SYBR Green I technology on the LightCycler was then performed. Master-mix for each PCR run was prepared as follows: 6.4 μl water, 1.2 μl $MgCl_2$ (4 mM), 0.2 μl of each primer (4 pmol), 1 μl Fast Start DNA Master SYBR Green I mix (Roche Diagnostics). Finally, 9 μl master-mix and 25 ng reverse transcribed total RNA in 1 μl water were transferred into capillaries, reaching end volume 10 μl. The following amplification program was used: after 10 min of denaturation at 95 °C, 40 cycles of real-time PCR with 3-segment amplification were performed consisting of 15 s at 95 °C for denaturation, 10 s at 60 °C for annealing and 20 s at 72 °C for polymerase elongation. The melting step was then performed with slow heating starting at 60 °C with a rate of 0.1 °C per second up to 99 °C with continuous measurement of fluorescence. The expressions of the UBQ, GAPD, β-actin and 18S rRNA were quantified separately. Further on, 10 target genes (TG) of interest were amplified: IGF-1, IGF-2, IGFR-1, IGFR-2, IGFBP-1 to IGFBP-6. These factors, all members of the somatotropic axis, were supposed to vary during the *Estrumate* treatment. In each biological sample all 14 mRNA transcripts were quantified.

Data acquisition

Data on the expression levels of studied factors were obtained in the form of crossing points (CP) as described earlier (Rasmussen 2001). The data acquisition was done employing the '*second derivative maximum*' method (Rasmussen 2001) as computed by the LightCycler Software 3.5 (Roche Diagnostics). For further data analysis the Excel based application *BestKeeper* was programmed to accelerate the computing procedure.

Table 1. Descriptive statistics of four candidate housekeeping genes (HKG) based on their crossing point (CP) values. In the two last columns the *BestKeeper* index is computed together with the same descriptive parameters, either for four genes (UBQ, GAPD, β-actin and 18S) or for three genes after removal of 18S (UBQ, GAPD and β-actin).

Data of candidate housekeeping genes ($n = 4$)						
Factor	UBQ	GAPD	β-actin	18S	*BestKeeper* ($n = 4$)	*BestKeeper* ($n = 3$)
N	31	31	31	31	31	31
GM [CP]	20.83	21.48	18.26	12.83	17.99	20.14
AM [CP]	20.86	21.5	18.29	12.97	18.03	20.16
Min [CP]	19.22	19.65	16.71	9.87	16.44	18.65
Max [CP]	23.19	24.3	20.8	16.58	20.86	22.65
SD [± CP]	0.76	0.74	0.79	1.5	0.9	0.69
CV [% CP]	3.66	3.45	4.34	11.57	4.98	3.43
Min [x-fold]	−3.06	−3.56	−2.93	−7.81	2.93	2.8
Max [x-fold]	5.13	7.05	5.82	13.44	7.31	5.7
SD [± x-fold]	±1.7	±1.67	±1.73	±2.83	±1.86	±1.61

Abbreviations: N: number of samples; GM [CP]: the geometric mean of CP; AM [CP]: the arithmetic mean of CP; Min [CP] and Max [CP]: the extreme values of CP; SD [± CP]: the standard deviation of the CP; CV [% CP]: the coefficient of variance expressed as a percentage on the CP level; Min [x-fold] and Max [x-fold]: the extreme values of expression levels expressed as an absolute x-fold over- or under-regulation coefficient; SD [± x-fold]: standard deviation of the absolute regulation coefficients.

Analysis of expression stability of housekeeping genes

Descriptive statistics of the derived crossing points were computed for each HKG: the geometric mean (GM), arithmetic mean (AM), minimal (Min) and maximal (Max) value, standard deviation (SD), and coefficient of variance (CV). All CP data are compared over the entire study, including control and all treatment groups. Herein, four genes, each of $n = 31$, were investigated. The x-fold over- or under-expression of individual samples towards the geometric mean CP are calculated and the multiple factor of their minimal and maximal values, expressed as the x-fold ratio and its standard deviation, are presented [Equations (1) and (2), Table 1]. These x-fold regulation results are corrected via the factor specific real-time PCR efficiency, calculated according Equation (3).

$$Min[x\_fold] = \boldsymbol{E}^{\min[CP]-GM[CP],} \quad (1)$$

$$Max[x\_fold] = \boldsymbol{E}^{\max[CP]-GM[CP].} \quad (2)$$

The corresponding real-time PCR efficiency **(E)** can be obtained in two ways. It can be computed either as sample specific (Tichopad *et al.* 2003, Liu & Saint 2002), or as factor specific (Rasmussen 2000) according to Equation (3). The slope of linear regression model fitted over log-transformed data of serially diluted input DNA concentrations plotted against their CPs (Rasmussen 2000, Pfaffl 2001). The maximal efficiency of PCR is $\mathbf{E = 2}$ where every single template is replicated in each cycle and the minimal value is $\mathbf{E = 1}$, corresponding to no replication.

$$\boldsymbol{E = 10}^{-1/\text{slope.}} \quad (3)$$

After the descriptive statistics for the individual candidate, HKG expression levels have been calculated, the first estimation of HKG expression stability can already be done, based on the inspection of calculated variations (SD and CV values). According to the variability observed, HKGs can be ordered from the most stably expressed, exhibiting the lowest variation, to the least stable one, exhibiting the highest variation. Any studied gene with the SD higher than 1 (= starting template variation by the factor 2) can be considered inconsistent (Table 1).

From the genes considered stably expressed, the *BestKeeper Index* specific for the respective sample is calculated as the geometric mean (3) of its candidate HKGs CP values [Equation (4)], where z is the total number of HKGs included.

$$BestKeeper\ Index = \sqrt[z]{CP_1 \times CP_2 \times CP_3 \times ..... \times CP_z.} \quad (4)$$

Table 2. Repeated pair-wise correlation analysis and correlation analysis of candidate housekeeping genes (HKG). A: Genes are pair-wise correlated one with another and then with the *BestKeeper* index ($n = 4$); B: results of the correlation analysis HKG versus *BestKeeper* index is shown ($n = 3$).

2A: Repeated pair-wise correlation analysis ($n = 4$)				
vs.	HKG 1 UBQ	HKG 2 GAPD	HKG 3 β-actin	HKG 4 18S
HKG 2	0.771	–	–	–
p-Value	0.001	–	–	–
HKG 3	0.728	0.803	–	–
p-Value	0.001	0.001	–	–
HKG 4	0.486	0.554	0.576	–
p-Value	0.006	0.001	0.001	–
BestKeeper vs.	UBQ	GAPD	β-actin	18S
Coeff. of corr. [r]	0.766	0.823	0.832	0.902
p-Value	0.001	0.001	0.001	0.001
Repeated pair-wise correlation analysis ($n = 4$) HKG vs. *BestKeeper* index out of 4				
HKG	HKG 1 UBQ	HKG 2 GAPD	HKG 3 β-actin	HKG 4 18S
Coeff. of corr. [r]	0.766	0.823	0.832	0.902
Coeff. of det. [r^2]	0.587	0.677	0.692	0.814
p-Value	0.001	0.001	0.001	0.001
2B: Repeated pair-wise correlation analysis ($n = 3$)				
vs.	HKG 1 UBQ	HKG 2 GAPD	HKG 3 $\beta-$actin	HKG 4
HKG 2	0.771	–	–	–
p-Value	0.001	–	–	–
HKG 3	0.728	0.803	–	–
p-Value	0.001	0.001	–	–
HKG 4	–	–	–	–
p-Value	–	–	–	–
BestKeeper vs.	UBQ	GAPD	β-actin	
Coeff. of corr. [r]	0.903	0.929	0.926	
p-Value	0.001	0.001	0.001	
Repeated pair-wise correlation analysis ($n = 3$) HKG vs. *BestKeeper* index out of 3				
HKG	HKG 1 UBQ	HKG 2 GAPD	HKG 3 β-actin	HKG 4
Coeff. of corr. [r]	0.903	0.929	0.926	–
Coeff. of det. [r^2]	0.815	0.863	0.857	–
p-Value	0.001	0.001	0.001	–

Analysis of the inter-HKG relations

To estimate inter-gene relations of all possible HKG pairs, numerous *pair-wise correlation analyses* are performed. Within each such correlation the *Pearson correlation coefficient* (r) and the probability p value are calculated (Tables 2A and 2B). All those highly correlated HKGs are combined into an index. Then, correlation between each candidate HKG and the index is calculated, describing the relation between the index and the contributing candidate HKG by the Pearson correlation coefficient (r), coefficient of determination (r^2) and the p-value (Tables 2A and 2B).

Analysis of target genes

Target gene (TG) expression data are statistically processed in the same way like those of HKGs, e.g., their GM, AM, SD, CV, Min. and Max. values (Table 4). Also here the *pair-wise correlation analyses* are performed to see any relation between pairs of TGs (Table 3).

To consider if a TG exhibits an expression pattern comparable or different from another TG, they are inspected in the same way as described for the HKGs and finally also correlated with the calculated index. Then, the same parameters of the correlation analysis as for HKG are calculated (Tables 4 and 5). Where a high correlation of TG to the index occurs, an expression pattern comparable to the HKG can be assumed. TGs expressed differentially from the index show no significance and sometimes even inverse correlation coefficients.

Analysis of sample integrity and expression stability within HKGs

Since the occurrence of outliers among prepared samples can obscure the accuracy of the estimation, individual samples are tested (herein $n = 31$) for their integrity (e.g. mRNA respectively cDNA quantity and quality) as well as their expression stability. An intrinsic variance (InVar) of expression for a single sample is calculated as a mean value square difference of single sample's CP value for one factor from a mean CP value of the same factor [Equation (5)].

$$\boldsymbol{InVar}_m[\pm\mathrm{CP}] = \frac{1}{n-1}\sum_{i=1}^{n}\left(CP_n^m - meanCP_n\right)^2, \quad (5)$$

Table 3. Descriptive statistics of target genes. Ten genes are analysed based on their CP values in the same way like HKGs (legend in Table 1).

Data of target genes ($n = 10$)										
Factor	TG 1 IGF-1	TG 2 IGF-2	TG 3 IGF-R-1	TG 4 IGF-R-2	TG 5 BP-1	TG 6 BP-2	TG 7 BP-3	TG 8 BP-4	TG 9 BP-5	TG 10 BP-6
N	31	31	31	31	31	31	31	31	31	31
GM [CP]	29.29	23.12	24.56	37.88	29.23	30.51	29.95	31.09	26.7	30.32
AM [CP]	29.31	23.14	24.59	37.89	29.38	30.53	30	31.13	26.74	30.36
Min [CP]	27.59	21.54	23.17	36.54	24.59	28.47	27.13	28.88	23.52	27
Max [CP]	31.42	25.52	27.68	39.92	35.33	33.09	36.47	34.41	29.66	33.52
SD [± CP]	0.79	0.86	0.88	0.66	2.49	0.77	1.32	1.12	1.25	1.1
CV [% CP]	2.71	3.71	3.59	1.74	8.47	2.51	4.41	3.59	4.68	3.64
Min [x-fold]	−3.26	−2.99	−2.63	−2.54	−24.92	−4.12	−7.06	−4.64	−9.06	−10.02
Max [x-fold]	4.37	5.29	8.67	4.1	68.62	5.96	91.86	9.96	7.78	9.16
SD [± x-fold]	1.73	1.81	1.84	1.58	5.61	1.7	2.5	2.17	2.38	2.15

Table 4. Pair-wise correlation analysis of the ten target genes. Target genes are pair-wise correlated among each other. Pearson correlation coefficient (r) and the value of probability p are shown.

Repeated pair-wise correlation analysis [Pearson correlation coefficient (r)]										
vs.	IGF-1 TG 1	IGF-2 TG 2	IGF-R-1 TG 3	IGF-R-2 TG 4	BP-1 TG 5	BP-2 TG 6	BP-3 TG 7	BP-4 TG 8	BP-5 TG 9	BP-6 TG 10
TG 2	0.367	–	–	–	–	–	–	–	–	–
p-Value	0.043	–	–	–	–	–	–	–	–	–
TG 3	0.43	0.586	–	–	–	–	–	–	–	–
p-Value	0.016	0.001	–	–	–	–	–	–	–	–
TG 4	0.073	−0.03	−0.068	–	–	–	–	–	–	–
p-Value	0.699	0.874	0.714	–	–	–	–	–	–	–
TG 5	−0.003	−0.176	0.345	0.064	–	–	–	–	–	–
p-Value	0.984	0.345	0.057	0.729	–	–	–	–	–	–
TG 6	0.257	0.331	0.309	0.102	−0.019	–	–	–	–	–
p-Value	0.163	0.069	0.091	0.587	0.921	–	–	–	–	–
TG 7	0.252	0.612	0.81	−0.006	0.377	0.189	–	–	–	–
p-Value	0.172	0.001	0.001	0.976	0.037	0.307	–	–	–	–
TG 8	0.257	0.832	0.711	0.109	0.057	0.291	0.738	–	–	–
p-Value	0.163	0.001	0.001	0.56	0.759	0.112	0.001	–	–	–
TG 9	0.044	−0.232	0.054	0.269	0.139	0.321	−0.056	0.016	–	–
p-Value	0.812	0.211	0.774	0.144	0.453	0.078	0.766	0.929	–	–
TG 10	0.335	0.379	0.283	0.174	−0.123	0.563	0.116	0.425	0.441	–
p-Value	0.066	0.035	0.123	0.35	0.508	0.001	0.534	0.017	0.013	–
BestKeeper vs.	TG 1	TG 2	TG 3	TG 4	TG 5	TG 6	TG 7	TG 8	TG 9	TG 10
Coeff. of corr. [r]	0.402	0.775	0.665	0.192	−0.041	0.18	0.696	0.811	−0.132	0.266
p-Value	0.025	0.001	0.001	0.302	0.827	0.33	0.001	0.001	0.477	0.147

Table 5. Results of pair-wise correlation analysis of target gene vs. *BestKeeper* index.

Repeated pair-wise correlation analysis: TG vs. *BestKeeper* ($n = 3$ HKG)										
	TG 1 IGF-1	TG 2 IGF-2	TG 3 IGF-R-1	TG 4 IGF-R-2	TG 5 BP-1	TG 6 BP-2	TG 7 BP-3	TG 8 BP-4	TG 9 BP-5	TG 10 BP-6
Coeff. of corr. [r]	0.4	0.78	0.67	0.19	−0.04	0.18	0.7	0.81	−0.13	0.27
Coeff. of det. [r^2]	0.16	0.6	0.44	0.04	0	0.03	0.48	0.66	0.02	0.07
p-Value	0.025	0.001	0.001	0.302	0.827	0.33	0.001	0.001	0.477	0.147

where the term in brackets denotes a difference of respective CP observation (n) of respective HKG (m) from the average CP value of the same HKG. Results are expressed in CP units [± CP] or as percentage of the mean [± %CP]. Further, it is expressed as an efficiency corrected intrinsic variation of x-fold, over- or under-expression of studied factor in the respective sample towards the mean CP of the same factor [± x-fold] [Equation (6)].

$$\boldsymbol{InVar}_m[\pm x\_fold] = \boldsymbol{E}_m^{InVar[\pm CP]}. \quad (6)$$

If justified, strongly deviating samples, due to inefficient sample preparation, incomplete reverse transcription or sample degradation, can be removed from the *BestKeeper* index calculation and its consistence and reliability thus be increased. A removal is recommended over a 3-fold over- or under-expression.

Results and discussion

In this paper, the Excel based tool *BestKeeper,* is presented and was tested in biological materials. The software is able to compare expression levels of up to ten HKGs together with ten TGs, each in up to hundred biological samples. Raw data input in the *BestKeeper* software are on Excel tables, separate for HKGs and TGs. Calculation proceeds in the background and results obtained can be easily printed out. All CP data are plotted in Excel table attached figures. It determines the 'optimal' HKGs employing the *pair-wise correlation analysis* of all pairs of candidate genes and calculates the geometric mean of the 'best' suited ones. The weighted index is correlated with up to ten target genes using the same pair-wise correlation analysis. Data observations are in form of raw CP (Rasmussen 2001) or threshold cycles (Ct) (Livak 2001) generated by a real-time PCR platform. The raw CPs seem to be best estimators of the expression levels as they are (in most cases) normally distributed and a parametric test can thus be performed. Expression data phrased in CP units is comparable with a logarithmic data transformation to the basis of two. This also gives the CP datasets the *Gaussian* distribution justifying usage of parametric methods.

Heterogeneous variance between groups of differently expressed genes, however, invalidates the use of *Pearson correlation coefficient*. Low expressed genes where CPs were obtained somewhere around cycles 30–35 surely show different variance compared to high expressed genes with CPs around 15 or even less. Such two samples cannot be correlated parametrically but on their ranks only. New version of the *BestKeeper* tool is, being prepared, employing also non-parametric methods such as the *Spearman* and *Kendall Tau correlation coefficient*. These methods are useful where genes with very different expression levels are compared.

Herein the software tool was tested on experimental data obtained from total RNA samples extracted from bovine *corpora lutea* under the *Estrumate* treatment. Compared to UBQ, GAPD and β-actin, in 18S, high CP variation in the expression was observed – a reason to exclude 18S from index calculation. On the other hand, all four HKG correlated very well one with another – a reason to retain 18S in the index. Both alternatives were tested and the correlation matrix for four candidate genes are shown in the Tables 2A and 2B. The expressions of UBQ, GAPD and β-actin showed CP variations around 0.75 CP (0.74 CP $<$ SD $<$ 0.79 CP), whereas the 18S expression showed high CP variation (SD = 1.5 CP) as well as up-/down-regulation (± 2.83-fold). Therefore the weighted index, calculated out of 4 candidates, showed a SD = 0.90 cycles. After the exclusion of 18S from index its variation decreased (SD = 0.69 cycles). The analysis showed a strong correlation ($0.766 < r < 0.902$) for all candidates.

Good consistence of the index was proved as its contributing housekeeping genes were tightly correlated with it. In both trials (with and without 18S) a good correlation with high significance level ($p < 0.001$) was observed, but after 18S removal, the significance increased (only rounded data are shown) and the correlation between the remaining HKGs and the index increased ($0.903 < r < 0.929$).

In above-shown way, a robust standardising index based on three HKGs was defined for a gene expression studies on bovine *corpora lutea*. Three genes represent a realistic calculation basis in a common laboratory and the minimal necessary number for a good performance of the analysis.

Correlation analyses of target genes showed (Table 3) that there were some significantly correlated genes (e.g. IGFBP-3 vs. IGFBP-4 and IGF-R-1 vs. IGFBP-4). Similarly, some target genes such as IGF-2, IGF-R-1, IGFBP-3 and BP-4 showed high correlation with the *BestKeeper* index. Tight correlation between applied internal standard and target gene shows regulation stability similar to the standard. Such a target gene can possibly be incorporated into the index.

Numerous genes were differentially expressed in this study, as they were not significantly correlated with the index (e.g. IGF-1, IGF-R-2, IGFBP-1, IGFBP-2, IGFBP-5, IGFBP-6). Some genes exhibited even totally inverse regulation of the expression, e.g. IGFBP-1 and IGFBP-5 as reflected by the negative correlation index (Tables 4 and 5).

Sample integrity was investigated using all four HKGs (no data shown). The InVar of the investigated 31 samples had low CP variation as well as on x-fold level. Three of the investigated samples showed higher variations in the expression stability of the HKGs, but still in the range of acceptance within a 3-fold regulation.

The earlier presented *GeNorm* software (Vandesompele *et al.* 2002) is restricted to the HKG analysis only, whereas, in *BestKeeper* software, additionally up to ten TGs can be analysed. Once a robust *BestKeeper* index was constructed, it can be applied as an expression standard in the same way like any single housekeeping gene. For a subsequent data processing, the CP datasets can be imported into analysis software tools such as *REST* (Pfaffl *et al.* 2002), *GeNorm* (Vandesompele *et al.* 2002) or *Q-Gene* (Muller *et al.* 2002). The *BestKeeper* application and necessary information about data processing and handling can be downloaded on http://www.wzw.tum.de/gene-quantification/bestkeeper.html

References

Chomczynski PA (1993) Reagent for single-step simultaneous isolation of RNA. *BioTechniques* **15**: 532–536.

Foss DL, Baarsch MJ, Murtaugh MP (1998) Regulation of hypoxanthine phosphoribosyltransferase, glyceraldehyde-3-phosphate dehydrogenase and beta-actin mRNA expression in porcine immune cells and tissues. *Anim. Biotechnol*. **9**: 67–78.

Liu W, Saint DA (2002) A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Anal. Biochem*. **302**: 52–59.

Livak KJ (1997 & 2001) ABI Prism 7700 Sequence Detection System User Bulletin #2 Relative Quantification of Gene Expression.

Muller PY, Janovjak H, Miserez R, Dobbie Z (2002) Processing of gene expression data generated by quantitative real-time RT-PCR. *BioTechniques* **32**: 2–7.

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucl. Acids Res.* **29**: e45.

Pfaffl MW, Hageleit M (2001) Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT-PCR. *Biotechnol. Lett.* **23**: 275–282.

Pfaffl MW, Horgan GW, Dempfle L (2002) Relative Expression Software Tool (REST©) for group wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucl. Acids Res.* **30**: e36.

Pfaffl MW, Mircheva Georgieva T, Penchev Georgiev I, Ontsouka E, Hageleit M, Blum JW (2002) Real-time RT-PCR quantification of insulin-like growth factor (IGF)-1, IGF-1 receptor, IGF-2, IGF-2 receptor, insulin receptor, growth hormone receptor, IGF-binding proteins 1, 2 and 3 in the bovine species. *Domest. Anim. Endocrinol.* **22**: 91–102.

Rasmussen R (2001) Quantification on the LightCycler instrument. In: Meuer S, Wittwer C, Nakagawara K, eds. *Rapid Cycle Real-Time PCR: Methods and Applications.* Heidelberg: Springer-Verlag Press, pp. 21–34.

Schmittgen TD, Zakrajsek BA (2000) Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *J. Biochem. Biophys. Meth.* **46**: 69–81.

Schuchhardt J, Beule D, Malik A, Wolski E, Eickhoff H, Lehrach H, Herzel H (2000) Normalisation strategies for cDNA microarrays. *Nucl. Acids Res.* **28**: e47.

Serazin-Leroy V, Denis-Henriot D, Morot M, de Mazancourt P, Giudicelli Y (1998) Semi-quantitative RT-PCR for comparison of mRNAs in cells with different amounts of housekeeping gene transcripts. *Mol. Cell. Probes* **12**: 283–291.

Suzuki T, Higgins PJ, Crawford DR (2000) Control selection for RNA quantitation. *BioTechniques* **29**: 332–337.

Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, Grisar T, Igout A, Heinen E (1999) Housekeeping genes as internal standards: use and limits. *J. Biotechnol*. **75**: 291–295.

Tichopad A, Dilger M, Schwarz G, Pfaffl MW (2003) Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucl. Acids Res*. **31**: e122.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalisation of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Gen. Biol.* **3**: 1–12.

Wittwer CT, Ririe KM, Andrew RV, David DA, Gundry RA, Balis UJ (1997) The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques* **22**: 176–181.

ELSEVIER

Molecular and Cellular Probes 18 (2004) 45–50

MOLECULAR AND CELLULAR PROBES

www.elsevier.com/locate/ymcpr

Inhibition of real-time RT–PCR quantification due to tissue-specific contaminants

Ales Tichopad, Andrea Didier, Michael W. Pfaffl*

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

Received 5 March 2003; accepted for publication 5 September 2003

Abstract

Real-time reverse transcription–polymerase chain reaction (RT–PCR) is currently considered the most sensitive method to study low abundance gene expression. Since comparison of gene expression levels in various tissues is often the purpose of an experiment, we studied a tissue-linked effect on nucleic acid amplification. Based on the raw data generated by a LightCycler instrument, we propose a descriptive mathematical model of PCR amplification. This model allowed us to study amplification kinetics of four common housekeeping genes in total RNA samples derived from various bovine tissues. We observed that unknown tissue-specific factors can influence amplification kinetics but this affect can be ameliorated, in part, by appropriate primer selection.

Keywords: Quantitative polymerase chain reaction; Real-time reverse transcription–polymerase chain reaction; Gene expression; Housekeeping genes; Ubiquitin; β-actin; GAPDH; 18S rRNA

1. Introduction

Reverse transcription–polymerase chain reaction (RT–PCR) is the method of choice for quantifying low abundant mRNAs in material such as cells and tissues [1–4]. This method is fast and highly reproducible. Further, its high sensitivity is its principal advantage over other techniques.

In real-time PCR the quantification takes place within an exponential phase of the amplification curve [5]. A crossing point (CP) or threshold cycle (Ct) is then extrapolated to determine a starting amount of template molecules. The CP gives the researcher the first raw information about the expression level of a given gene.

All methods of gene quantification report their findings relative to a measurable base (e.g. copies per cell, weight of tissue, volume of blood, etc.). The correct choice of the denominator depends on the question asked and can significantly affect the quality of the results [6]. To obtain an actual number of copies, various 'absolute' standards are often employed [7–9], but even in these cases, the quantification is always relative as some errors in a protocol are inevitably present [6,10]. So called housekeeping or maintenance genes [11] such as actins, tubulins, albumins, ubiquitin, glyceraldehyd-3-phosphate dehydrogenase (GAPDH), 18S or 28S ribosomal subunits (rRNA) are often used as relative standards [12]. These genes are believed to undergo little, if any, variation in expression under most experimental treatments. Yet, there have been many reports on the regulation of these genes [12–14].

Another important criterion for reliable measurement and comparison of more than one gene is that all of the genes amplify equally. Experiments using normalization with housekeeping genes often overlook this parameter despite the fact that corrections have already be suggested in the literature [15–19].

Many factors present in samples as well as exogenous contaminants have been shown to inhibit PCR (review in Refs. [20,21]). For example, the presence of hemoglobin, fat, glycogen, cell constituents, Ca^{2+}, DNA or RNA concentration, and DNA binding proteins are important factors [20,21]. Additionally, exogenous contaminants such as glove powder and phenolic compounds from

Abbreviations: RT–PCR, reverse transcription–polymerase chain reaction; CP, crossing point; GAPDH, glyceraldehyd-3-phosphate dehydrogenase; FDM, first derivative maximum; SDM, second derivative maximum.

* Corresponding author. Tel.: +49-8161-71-3511; fax: +49-8161-71-4204.

E-mail address: pfaffl@wzw.tum.de (M.W. Pfaffl).

0890-8508/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mcp.2003.09.001

the extraction process or the plastic ware can have an inhibiting effect [20,21].

Since some experiments compare gene expression in different organs [9,22], tissue-specific inhibition of DNA amplification may be important. To study the amplification inhibition associated with three randomly chosen tissue types we proposed a mathematical model describing the DNA amplification kinetics in real-time PCR. Using this model we could compare parameters of the amplification kinetics and analyze them statistically.

2. Materials and methods

2.1. Preparation of cDNA samples

Samples of cerebellum, muscle and liver were gathered from six slaughtered cows, immediately frozen in liquid nitrogen and then stored at −80 °C until the total RNA extraction procedure was performed.

Tissue samples were homogenized and total RNA was extracted with a commercially available product, peqGOLD TriFast (Peqlab, Erlangen, Germany), utilizing a single modified liquid separation procedure [23]. No additional purification was performed. Constant amounts of 1000 ng of RNA were reverse-transcribed to cDNA using 200 units of MMLV Reverse Transcriptase (Promega, Mannheim, Germany) according to the manufacturers instructions.

Integrity of the DNA was determined by electrophoresis on 1% agarose gels. Nucleic acid concentrations were measured on a spectrophotometer (BioPhotometer, Eppendorf, Hamburg, Germany) at $OD_{260\ nm}$ with 220–1600 nm UVettes (Eppendorf). Purity of the RNA extracted was determined as the $OD_{260\ nm}/OD_{280\ nm}$ ratio with expected values between 1.8 and 2.0 (BioPhotometer). A possible trend between the samples and their $OD_{260\ nm}/OD_{280\ nm}$ values was examined.

2.2. Real-time PCR fluorescence data acquisition

Primer sequences of four common housekeeping genes; ubiquitin, β-actin, GAPDH and 18S rRNA were designed to span at least one intron (except for 18S rRNA) and synthesized commercially (MWG Biotech, Ebersberg, Germany) as shown in Table 1. PCR conditions were optimized on a gradient cycler (T-Gradient, Biometra, Göttingen, Germany) and subsequently on a LightCycler (Roche Diagnostic, Mannheim, Germany) [24] by analyzing the melting curves of the products [25]. Real-time PCR using SYBR Green I technology [26] on the LightCycler was then carried out to amplify cDNAs from the tissue samples.

Master-mix for each PCR run was prepared as follows: 6.4 μl of water, 1.2 μl $MgCl_2$ (4 mM), 0.2 μl of each primer (4 pmol), 1.0 μl Fast Start DNA Master SYBR Green I mix (Roche Diagnostics). Finally, 9 μl of master-mix and 25 ng of reverse transcribed total RNA in 1 μl water were transferred into capillaries (end volume 10 μl).

The following amplification program was used: After 10 min of denaturation at 95 °C, 40 cycles of real-time PCR with three-segment amplification were performed with: 15 s at 95 °C for denaturation, 10 s at respective annealing temperature (Table 1) and 20 s at 72 °C for elongation. A melting step was then performed with slow heating starting at 60 °C with a rate of 0.1 °C/s up to 99 °C with continuous measurement of fluorescence. The same gene was always quantified in each run to prevent any inter-run variation.

Fluorescence data from real-time PCR experiments were taken directly from LightCycler software version 3 (Roche Diagnostics), exported to SigmaPlot 2000 (SPSS, Munich, Germany) and fitted with a '*Four-parametric sigmoid model*' as described earlier by our group [27]. Parameters a, b, x_0 and y_0 of each fit were documented together with the coefficient of determination r^2.

All statistics were done in SigmaPlot 2000 (SPSS) and SigmaStat 2.0 (SPSS, Jandel Corporation).

2.3. Crossing point (CP) acquisition

On each individual real-time PCR run, five different CPs were acquired based on different determination procedures. First, the CP was placed into the first derivative maximum ($FDM_{SM} = x_0$) and into the second derivative maximum of the four-parametric sigmoid model (SDM_{SM}) of each run as shown earlier [27].

Table 1
Details of primers used to amplify four housekeeping genes

Gene	Primers	Sequence length (bp)	Annealing temperature (°C)
Ubiquitin	for: AGA TTC AGG ATA AGG AAG GCA T rev: GCT CCA CCT CCA GGG TGA T	198	60
GAPDH	for: GTC TTC ACT ACC ATG GAG AAG G rev: TCA TGG ATG ACC TTG GCC AG	197	58
18S rRNA	for: GAG AAA CGG CTA CCA CAT CCA A rev: GAC ACT CAG CTA AGA GCA TCG A	338	60
β-actin	for: AAC TCC ATC ATG AAG TGT GAC G rev: GAT CCA CAT CTG CTG GAA GG	234	60

Table 2
Two-way ANOVA

Factor	a	b	FDM_{SM}	SDM_{SM}	FP_{LC}	SDM_{LC}	CP_{Tm}
Tissue	0.01	<0.001	0.004	0.02	0.005	0.008	0.004
Gene	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Tissue–gene interaction	0.004	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

P-values of significance. Each of three rows indicates either one of factors or their interaction. In columns, *P*-values of effect of factors (or interaction) on respective parameter are shown.

Further, CP was computed using the '*Fit point method*' (FP_{LC}) [5] and '*Second derivative maximum method*' (SDM_{LC}) [5,28], both part of the LightCycler software 3.3 (Roche Diagnostics). In the FP_{LC} method, uninformative background fluorescence observations were discarded by setting a constant noise band. An intersecting line was then arbitrarily placed at the base of the exponential portion of the amplification curves. This generated CPs acquired at a constant fluorescence level (value 2 in our case).

In the SDM_{LC} method the second derivative maximum is calculated by LightCycler software based on an unknown and unpublished mathematical approximation of partial amplification kinetics around the supposed SDM_{LC} [5,28]. The FP_{LC} and SDM_{LC} were directly obtained from the calculated values by the LightCycler software 3.3 (Roche Diagnostics).

Eventually, the '*Taqman threshold level*' (Ct) or CP [29] computing method was simulated by fitting the intersecting line upon the 10 times value of ground fluorescence standard deviation (CP_{Tm}). In the '*Taqman threshold level*' procedure, the y_0 values of the four-parametric sigmoid model were considered ground fluorescence.

While parameters a and b describe amplification kinetics, FDM_{SM}, SDM_{SM}, FP_{LC}, SDM_{LC}, and CP_{Tm} are considered quantification parameters since they are clearly defined constants within the model.

Table 3a
Statistically processed parameters a, b, FDM_{SM}, SDM_{SM}, FP_{LC}, SDM_{LC}, CP_{Tm}, and r^2 of ubiquitin amplification

Tissue		a	b	FDM_{SM}	SDM_{SM}	FP_{LC}	SDM_{LC}	CP_{Tm}	r^2
Cerebellum	Mean	43.118	1.950	25.649	23.082	20.180	21.817	22.680	1.000
	CV (%)	9.56	1.17	1.31	1.50	2.13	1.43	1.77	0.004
Liver	Mean	39.355	2.004	26.184	23.545	20.688	22.288	22.597	1.000
	CV (%)	7.79	1.62	1.43	1.64	1.73	1.79	1.47	0.010
Muscle	Mean	41.958	2.064	26.443	23.725	20.637	22.487	25.370	0.999
	CV (%)	5.40	2.25	0.81	0.94	1.52	1.15	0.67	0.018
$Mean_{total}$		41.477	2.006	26.092	23.450	20.502	22.197	23.549	1.000
$CV_{in\text{-}tissue}$ (%)		7.58	1.68	1.18	1.36	1.79	1.46	1.30	0.011
$CV_{out\text{-}tissue}$ (%)		4.65	2.85	1.55	1.41	1.36	1.55	6.70	0.014

P-values of significance. Each of three rows indicates either one of factors or their interaction. In columns, *P*-values of effect of factors (or interaction) on respective parameter are shown.

Table 3b
Statistically processed parameters a, b, FDM_{SM}, SDM_{SM}, FP_{LC}, SDM_{LC}, CP_{Tm}, and r^2 of GAPDH amplification

Tissue		a	b	FDM_{SM}	SDM_{SM}	FP_{LC}	SDM_{LC}	CP_{Tm}	r^2
Cerebellum	Mean	47.223	2.075	23.663	20.930	18.185	19.583	20.483	0.998
	CV (%)	11.48	1.43	1.14	1.36	1.90	1.43	2.01	0.009
Liver	Mean	46.675	2.094	24.936	22.179	19.322	20.868	21.580	0.998
	CV (%)	6.39	2.75	1.61	1.97	2.09	2.21	2.20	0.020
Muscle	Mean	52.415	2.228	21.588	18.653	15.800	17.440	16.377	0.997
	CV (%)	3.79	2.94	3.30	4.08	4.70	4.06	4.48	0.032
$Mean_{total}$		48.771	2.132	23.396	20.587	17.769	19.297	19.480	0.998
$CV_{in\text{-}tissue}$ (%)		7.22	2.37	2.02	2.47	2.90	2.57	2.89	0.020
$CV_{out\text{-}tissue}$ (%)		6.50	3.92	7.22	8.68	10.12	8.98	14.08	0.068

P-values of significance. Each of three rows indicates either one of factors or their interaction. In columns, *P*-values of effect of factors (or interaction) on respective parameter are shown.

Table 3c
Statistically processed parameters a, b, FDM_{SM}, SDM_{SM}, FP_{LC}, SDM_{LC}, CP_{Tm}, and r^2 of 18S rRNA amplification

Tissue		a	b	FDM_{SM}	SDM_{SM}	FP_{LC}	SDM_{LC}	CP_{Tm}	r^2
Cerebellum	Mean	49.782	2.701	15.274	11.717	9.518	10.556	10.923	0.996
	CV (%)	3.76	5.33	3.64	6.26	6.32	5.83	6.17	0.047
Liver	Mean	53.544	2.897	14.669	10.854	8.638	9.809	9.185	0.996
	CV (%)	3.35	2.55	9.01	12.21	12.28	12.38	11.15	0.040
Muscle	Mean	55.943	2.752	15.369	11.744	9.250	10.573	10.267	0.997
	CV (%)	2.67	2.31	5.61	7.76	8.32	7.94	7.75	0.041
$Mean_{total}$		53.090	2.784	15.104	11.439	9.135	10.313	10.125	0.996
$CV_{in\text{-}tissue}$ (%)		3.26	3.40	6.09	8.74	8.97	8.72	8.36	0.042
$CV_{out\text{-}tissue}$ (%)		5.85	3.66	2.51	4.43	4.94	4.23	8.67	0.019

P-values of significance. Each of three rows indicates either one of factors or their interaction. In columns, *P*-values of effect of factors (or interaction) on respective parameter are shown.

2.4. Statistical evaluation of model parameters

Two-way ANOVA with tissue as the first factor of three levels (cerebellum, muscle and liver) and gene as the second factor of four levels (ubiquitin, β-actin, GAPDH, 18S rRNA) was applied to the parameters a, b, FDM_{SM}, SDM_{SM}, FP_{LC}, SDM_{LC} and CP_{Tm} (Table 2). Normal distribution was given within the data sets.

For all above-mentioned parameters and r^2 following statistical indicators were calculated (Tables 3a–3d)

- Interaction mean (i.e. from the six values within one level of factor gene and one level of factor tissue) and interaction coefficient of variance-CV.
- Total mean ($mean_{total}$) out of 18 values (always six samples in three tissues) for each factor gene.
- Mean value out of three CVs ($CV_{in\text{-}tissue}$) reporting internal variance within all three tissue levels.
- Coefficient of variance out of three interaction means ($CV_{out\text{-}tissue}$) showing a variability caused by factor tissue.

3. Results and discussion

All primers used could satisfactorily amplify the flanked sequence. The melting curve analysis and gel analysis detected very little, if any, nonspecific product. We approximated the PCR amplification kinetics with the four-parametric sigmoid model. This model describes well (in all data sets $r^2 > 0.99, n = 40$) the entire fluorescence curve and therefore its beginning and end do not need to be arbitrarily delimited [19]. Nevertheless, correlation between values of b and r^2 showed that there were differences in the goodness of the fit (Pearson correlation coefficient $r = 0.915, n = 72$). The best fit was in runs with high amplification efficiencies. With decreasing amplification efficiency the determination power of the model also decreased.

There is an integral purification step at the end of the extraction procedure [23], consisting of repeated washing the final total RNA pellet with ethanol. In this study no additional RNA purification was performed since additional purification decreases yield and is often omitted. This procedure simulated a routine PCR sample preparation as it is carried out in most labs. The contamination within the RNA samples detected as $OD_{260\ nm}/OD_{280\ nm}$ ratios was not significantly related to the type of tissue (data not shown).

Statistical analysis of the parameters a and b (Table 2) under an influence of the two experimental factors showed that the tissue was the largest source of variance and the primer sequences had the least affect [21,22].

A similar trend of variability within the log-linear trajectory slope (b) and plateau height (a) showed that the tissue from

Table 3d
Statistically processed parameters a, b, FDM_{SM}, SDM_{SM}, FP_{LC}, SDM_{LC}, CP_{Tm}, and r^2 of β-actin amplification

Tissue		a	b	FDM_{SM}	SDM_{SM}	FP_{LC}	SDM_{LC}	CP_{Tm}	r^2
Cerebellum	Mean	85.015	1.418	22.499	20.632	16.640	19.362	19.643	1.000
	CV (%)	5.11	2.15	2.22	2.54	3.21	2.69	2.34	0.004
Liver	Mean	86.694	1.467	23.555	21.624	17.400	20.348	18.633	1.000
	CV (%)	2.14	1.31	0.85	0.95	1.52	1.11	1.11	0.002
Muscle	Mean	84.886	1.470	24.264	22.328	18.230	21.047	20.813	1.000
	CV (%)	2.75	3.53	0.90	1.03	1.14	1.16	0.88	0.005
$Mean_{total}$		85.532	1.452	23.440	21.528	17.423	20.252	19.697	1.000
$CV_{in\text{-}tissue}$ (%)		1.00	2.33	1.32	1.51	1.96	1.65	1.45	0.004
$CV_{out\text{-}tissue}$ (%)		1.18	2.01	3.79	3.96	4.56	4.18	5.54	0.001

P-values of significance. Each of three rows indicates either one of factors or their interaction. In columns, *P*-values of effect of factors (or interaction) on respective parameter are shown.

which total RNA was extracted has a significant effect on the PCR kinetics and thus on the CP acquisition (Table 2). This can be caused by different amounts of cellular debris present in samples after RNA extraction [30,31]. Also endogenous contaminants such as blood or fat play an important role. Contamination of the sample may affect both the PCR as well as the preceding RT reaction [20,21].

Since interaction between both factors; tissue and gene is significant, the tissue-specific disturbance is not the same for all four amplified sequences but rather is sequence-specific. In our study, the highest resistance to tissue-specific disturbance showed the sequence of β-actin followed by ubiquitin, 18S rRNA and GAPDH (see $CV_{out\text{-}tissue}$ values in Tables 3a–3d). A plausible explanation of this interaction may be the presence of specific DNA blocking by polysaccharides or proteins present as endogenous contaminants in the sample [32]. It is possible that DNA amplification may be affected by regions of the template DNA that are specifically blocked by these endogenous macromolecules. Our data show that not only the choice of housekeeping genes [12–14] but also tissue-specific factors and the sequence-specific factors can affect the expression assays.

Tissue-specific suppression can be compensated, in part, by well performing primers such as those for β-actin and ubiquitin used here. From this data it seems that sequences that amplified with higher efficiency (i.e. small *b*) better resist inhibition and show lower variance in all parameters of the PCR kinetics (compare $mean_{total}$ of *b* and $CV_{out\text{-}tissue}$ values in Tables 3a and 3d with Tables 3b and 3c). Thus, primer selection and documenting the reaction efficiency are important PCR optimization steps. Although housekeeping genes are expressed differently in various tissues our data show that some vary less than others. For example, ubiquitin showed marginally higher variance between tissues than within one tissue (compare $CV_{out\text{-}group}$ with $CV_{in\text{-}group}$ in Table 3a). This suggests that the expression of ubiquitin in the different tissues was similar. The low variance for ubiquitin expression between tissues suggests that it is the best standard but is closely followed by β-actin and GAPDH. 18S rRNA, with its high variance, seems to be less suitable as an internal standard. This order was preserved in all CP computing methods.

Each method of computing CPs seems to be accurate for estimating expression levels but they varied slightly when CP acquisitions took place at different heights of the amplification curve (Tables 3a–3d). The method of first and second derivative maximum computed from the four-parametric sigmoid model is reliable and simple and generates reliable CPs comparable with other methods (see CV values in Tables 3a–3d).

Acknowledgements

The experimental animals were slaughtered at the EU-official slaughterhouse of the Bayerische Landesanstalt für Tierzucht at Grub, 85580 Poing, Germany.

References

[1] Schmittgen TD. Real-time quantitative PCR. Methods 2001;25:383–5.
[2] Orlando C, Pinzani P, Pazzagli M. Developments in quantitative PCR. Clinical Chemistry and Laboratory Medicin 1998;36:255–69.
[3] Gibson UE, Heid CA, Williams PM. A novel method for real time quantitative RT–PCR. Genome Research 1996;6:995–1001.
[4] Freeman WM, Walker SJ, Vrana KE. Quantitative RT–PCR: pitfalls and potential. BioTechniques 1999;26:112–25.
[5] Rasmussen R. Quantification on the lightcycler instrument. In: Meuer S, Wittwer C, Nakagawara K, editors. Rapid cycle real-time PCR: methods and applications. Heidelberg: Springer; 2001. p. 21–34.
[6] Ferré F. Quantitative or semi-quantitative PCR: reality versus myth. PCR Methods and Applications 1992;2:1–9.
[7] Pfaffl MW, Hageleit M. Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT–CR. Biotechnology Letters 2001;23:275–82.
[8] Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. Journal of Molecular Endocrinology 2000;25:169–93.
[9] Tichopad A, Pfaffl MW, Didier A. Tissue-specific expression pattern of bovine prion: quantification using real-time RT–PCR. Molecular and Cellular Probes 2003;17:5–10.
[10] Souazé F, Ntodou-Thomé A, Tran CY, Rostene W, Forgez P. Quantitative RT–PCR: limits and accuracy. BioTechniques 1996;21:280–5.
[11] Warrington JA, Nair A, Mahadevappa M, Tsygantskaya M. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiological Genomics 2000;2:143–7.
[12] Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, Grisar T, Igout A, Heinen E. Housekeeping genes as internal standards: use and limits. Journal of Biotechnology 1999;75:291–5.
[13] Schmittgen TD, Zakrajsek BA. Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT–PCR. Journal of Biochemical and Biophysical Methods 2000;20:69–81.
[14] Suzuki T, Higgins PJ, Crawford DR. Control selection for RNA quantitation. BioTechniques 2000;29:332–7.
[15] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)). Methods 2001;25:402–8.
[16] Pfaffl MW. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Research 2001;29(9):e45.
[17] Pfaffl MW, Horgan GW, Dempfle L. Relative Expression Software Tool (REST©) for group wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Research 2002;30(9):e36.
[18] Meijerink J, Mandigers C, van de Locht L, Tonnissen E, Goodsaid F, Raemaekers J. A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR. Journal of Molecular Diagnostics 2001;3:55–61.
[19] Weihong L, Saint A. A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. Analytical Biochemistry 2002;302:52–9.
[20] Wilson IG. Inhibition and facilitation of nucleic acid amplification. Applied and Environmental Microbiology 1997;63:3741–51.
[21] Rossen L, Norskov P, Holmstrom K, Rasmussen FO. Inhibition of PCR by components of food sample, microbial diagnostic assay and DNA-extraction solutions. International Journal of Food Microbiology 1992;17:37–45.
[22] Pfaffl MW, Lange IG, Daxenberger A, Meyer HH. Tissue-specific expression pattern of estrogen receptors (ER): quantification of ER

alpha and ER beta mRNA with real-time RT–PCR. Acta Pathologica Microbiologica et Immunologica Scandinavica 2001; 109:345–55.

[23] Chomczynski PA. Reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. BioTechniques 1993;15:532–4.

[24] Wittwer CT, Ririe KM, Andrew RV, David DA, Gundry RA, Balis UJ. The lightcycler: a microvolume multisample fluorimeter with rapid temperature control. BioTechniques 1997;22:176–81.

[25] Ririe KM, Rasmussen RT, Wittwer CT. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Analytical Biochemistry 1997;245:154–60.

[26] Morrison TB, Weis JJ, Wittwer CT. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. BioTechniques 1998;24:954–8.

[27] Tichopad A, Dzidic A, Pfaffl MW. Improving quantitative real-time RT–PCR reproducibility by boosting primer-linked amplification efficiency. Biotechnology Letters 2002;24:2053–6.

[28] Wittwer CT, Gutekunst M, Lohmann S. Method for quantification of an analyte. United States Patent No. US, 6,303,305 B1

[29] Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'–3' exonuclease activity of *Thermus aquaticus* DNA polymerase. Proceedings of the National Academy of Science of the United States of America 1991;15:7276–80.

[30] Wilson IG, Gilmour A, Cooper JE. Detection of toxigenic microorganisms in foods by PCR. In: Kroll RG, Gilmour A, Sussman M, editors. In new techniques in food and beverage microbiology. London: Blackwell; 1993. p. 163–72.

[31] Starnbach MN, Falkow S, Tomkins SL. Species-specific detection of *Legionella pneumophila* in water by DNA amplification and hybridization. Journal of Clinical Microbiology 1989;27:1257–61.

[32] Rijpens NP, Jannes G, Van Asbroeck M, Rossau R, Herman LMF. Direct detection of *Brucella* spp. In raw milk by PCR and reverse hybridization with 16S–23S rRNA spacer probes. Applied Environmental Microbiology 1996;62:1683–8.

Nucleic Acids Research, 2003, Vol. 31, No. 20 ***e122***
DOI: 10.1093/nar/gng122

Standardized determination of real-time PCR efficiency from a single reaction set-up

Ales Tichopad, Michael Dilger[1], Gerhard Schwarz[2] and Michael W. Pfaffl*

Institute of Physiology and [1]Institute of Agronomy and Plant Breeding, FML-Weihenstephan, Center of Life and Food Science, Technical University of Munich, Germany and [2]EpiGene GmbH, Biotechnology in Plant Protection, Hohenbachernstrasse 19–21, 85354 Freising, Germany

Received June 24, 2003; Revised July 29, 2003; Accepted August 25, 2003

ABSTRACT

We propose a computing method for the estimation of real-time PCR amplification efficiency. It is based on a statistic delimitation of the beginning of exponentially behaving observations in real-time PCR kinetics. PCR ground fluorescence phase, non-exponential and plateau phase were excluded from the calculation process by separate mathematical algorithms. We validated the method on experimental data on multiple targets obtained on the LightCycler platform. The developed method yields results of higher accuracy than the currently used method of serial dilutions for amplification efficiency estimation. The single reaction set-up estimation is sensitive to differences in starting concentrations of the target sequence in samples. Furthermore, it resists the subjective influence of researchers, and the estimation can therefore be fully instrumentalized.

INTRODUCTION

More than 10 years of PCR-based technologies have found their place in most of the laboratories involved in biomedical science. The application of PCR in gene expression studies is an example of a fast innovating field. So far, real-time PCR in combination with array techniques is the major approach adopted in quantitative gene expression studies. The fact that several nucleic acid molecules can be amplified up to microgram amounts opens the possibility to study gene regulation even in a single cell (1). The recent introduction of various fluorescence-based monitoring detection techniques into PCR (2–7) allowed the documentation of the amplification process in the so-called real-time PCR (8–10). The amplification of nucleic acids within the range of exponential growth of the reaction trajectory can be described by a pure exponential growth (equation **1**):

$$P = I * E^n \qquad \mathbf{1}$$

where P is the amount of the PCR product of the reaction, I is the input nucleic acid amount, E is the efficiency of the reaction ranging from 1 to 2 and n is the number of PCR cycles.

There is a constant tendency to place the quantification into an early phase of detectable amplification. In such an early portion of PCR trajectory the amplification has the exponential character described in equation **1**. The reaction trajectory at later reaction stages significantly diverges from the exponential type, and becomes a more stochastic process. In such an early portion of the amplification kinetics, a threshold fluorescence is set. As soon as the reaction reaches this threshold fluorescence, the information necessary for the quantitative judgment about the input concentration of the target sequence has been gathered (11). The fractional cycle number of threshold value (C_t) or crossing point (CP) is then compared with the CP of control samples. There are two ways of threshold level setting. It can be done either arbitrarily by using a randomly selected threshold or by applying computing algorithms. The maximum of the second or, generally, nth derivative of smoothed amplification kinetics gives a good and justified threshold level within the assay (11).

Since the result of a single quantitative PCR just reflects the relative amount of target sequence in the form of fluorescence units, it must be objectified by some control. Therefore, adequate quantitative information cannot be obtained from a real-time PCR assay unless at least two samples are analyzed. To make sure that RT reactions and amplification reactions proceed in a similar way in both samples, the amplification of another target sequence present in the sample is often introduced into the assay either simultaneously or in separated runs. The expression of the standard, called the housekeeping gene or reference gene, is assumed to be uninfluenced by experimental treatment and a similar detectable amplification product should therefore be obtained (12,13). Yet, there is a lot of evidence for regulation of these genes under defined treatments (14,15).

Recently, problems have been discussed, that different sequences were often amplified with different efficiencies, causing under/overestimation of input template copy numbers in orders of magnitude. The solution is to document the amplification efficiencies (E) of both reactions and to apply a compensating computing algorithm (16–18). The currently used and partly automated method of determination of amplification efficiency is the method of serial dilutions, each analyzed in triplicate (11). Using this method, serial dilutions of the starting template are prepared; in these, the input nucleic acid concentration is varied over several orders of magnitude. Usually, dilution series are prepared by serially

*To whom correspondence should be addressed. Tel: +49 8161 713511; Fax: +49 8161 714204; Email: pfaffl@wzw.tum.de

diluting the input nucleic acid five to 10 times with sterile water. Subsequently, the CP or C_t values are plotted against the log of the known starting concentration value and from the slope of the regression line the amplification efficiency is estimated (11,16). There are some variations of this method, but the serial dilution is always necessary. The method finally gives only one value of E for all dilution concentrations of the respective sequence. This is, however, a simplified approach, since the E varies considerably as the input concentration changes.

Therefore, what is required is a method of amplification efficiency determination that uses the reaction kinetics of a single sample. Since the amplification fluorescence raw data are available by data export from LightCycler (19) or ABI Prism Sequence Detection System (20) software, the efficiency estimation can be based on these data. Liu and Saint (21) suggested a method of amplification efficiency estimation based on absolute fluorescence increase in single reaction kinetics data. In this method, the portion of the data array believed to be exponentially behaving is taken, log-transformed and plotted. The authors consider the slope of the regression line the amplification efficiency. The idea behind this method is correct, but the crucial disadvantage consists of the researcher's subjective judgment; what data are exponentially behaving and what data are not. Furthermore, the necessary subjective delimitation procedure can not be instrumentalized. Delimitation of the exponential portion of the data is done precisely at the end of it, as the reaction kinetics here strongly depart from the exponential. A similar published method (22) is also based on the absolute fluorescence increase, but it takes place around the point of inflection of the quantification trajectory where a strong decaying trend of the amplification efficiency already occurs. This method is therefore underestimating the 'real efficiency'.

Here, we report a new method for a reliable estimation of real-time PCR efficiency, which is based on the fluorescence history of just a single reaction set-up and it resists any subjective manipulation. This method was applied on raw fluorescence data from the LightCycler real-time PCR platform.

MATERIALS AND METHODS

SRY plasmid DNA construction

The bovine SRY (male sex determining) gene coding sequence was cloned into pCR®4-TOPO® vector using the TOPO TA Cloning Kit for Sequencing (Invitrogen, Karlsruhe, Germany). This circular DNA construct was linearized with restriction digest and its purity was inspected on a 2% agarose gel. Exact quantification of the DNA content was done at $OD_{260\,nm}$ on a spectrophotometer (BioPhotometer®; Eppendorf, Hamburg, Germany) with UVettes (Eppendorf) in various dilutions and repeats (n = 12), to circumvent any source of error. For standard curve acquisition, six serial dilutions of linearized plasmid DNA ranging were prepared, representing 2.65×10^2–2.65×10^7 single-stranded (ss) SRY DNA molecules, serving as DNA templates for real-time PCR.

Real-time PCR on LightCycler

A primer pair flanking sequence within bovine SRY gene was constructed and synthesized (MWG Biotech, Ebersberg, Germany) as follows: forward primer, 5′-GAA CGC CTT CAT TGT GTG GTC-3′; reverse primer, 5′-TGG CTA GTA GTC TCT GTG CCT CCT-3′. The conditions for PCR were optimized in a gradient cycler (TGradient; Biometra, Göttingen, Germany) and subsequently in LightCycler (Roche Diagnostics, Mannheim, Germany) analyzing the melting curves of the products acquired (23). This was done with respect to primer annealing temperatures, primer concentrations, template concentrations and number of cycles applied. Real-time PCR using SYBR Green I technology (Roche Diagnostics) (10,19) with the above-mentioned primers was carried out amplifying cloned sequence in triplicate for each respective concentration. Master-mix for each PCR run was prepared as follows: 6.4 µl of water, 1.2 µl of $MgCl_2$ (4 mM), 0.2 µl of each primer (4 µM), 1.0 µl of Fast Start DNA Master SYBR Green I and 2.65×10^2–2.65×10^7 copies of ss SRY linearized plasmid DNA. The following amplification program was applied: after 10 min of denaturation at 95°C, 40 cycles of four-segment amplification were accomplished with: (i) 15 s at 95°C for denaturation, (ii) 10 s at 60°C for annealing, (iii) 20 s at 72°C for elongation and (iv) determination of fluorescence at an elevated temperature of 83°C (22). Subsequently, a melting curve program was applied with continuous fluorescence measurement.

RESULTS

After optimization of the real-time PCR assay with SRY, the gene sequence could be routinely run generating specific amplicons showing no primer dimers, a single sharp peak, identical melting points and an expected length of 164 bp in gel electrophoresis. The sensitivity of the LightCycler RT–PCR was evaluated using different starting amounts of cDNA in a standard curve. SYBR Green I fluorescence determination at the elevated temperature resulted in a reliable and sensitive cDNA quantification assay with high linearity (r = 0. 99) over six orders of magnitude from 2.65×10^2 to 2.65×10^7 recombinant standard DNA start molecules.

Determination of fluorescence ground phase in PCR

The earliest observation of detectable growth phase above the ground phase with sufficient n is well suited for estimation of E (Fig. 1, inlay).

To objectively detect the beginning of the exponential phase and to skip down the prior ground phase, a statistical method is applied. The ground phase is considered to behave linearly (equation **2**) and linear regression with intercept i_{lin} and slope β:

$$y_{lin} = i_{lin} + \beta \times x \qquad \mathbf{2}$$

Therefore, it can fit observations as long as there is no sudden significant increment of fluorescence due to reaction product generation. At the moment when the increment of fluorescence becomes a consistent trend, the beginning of the exponential phase takes place. To inspect whether each successively inspected observation still belongs to the linear

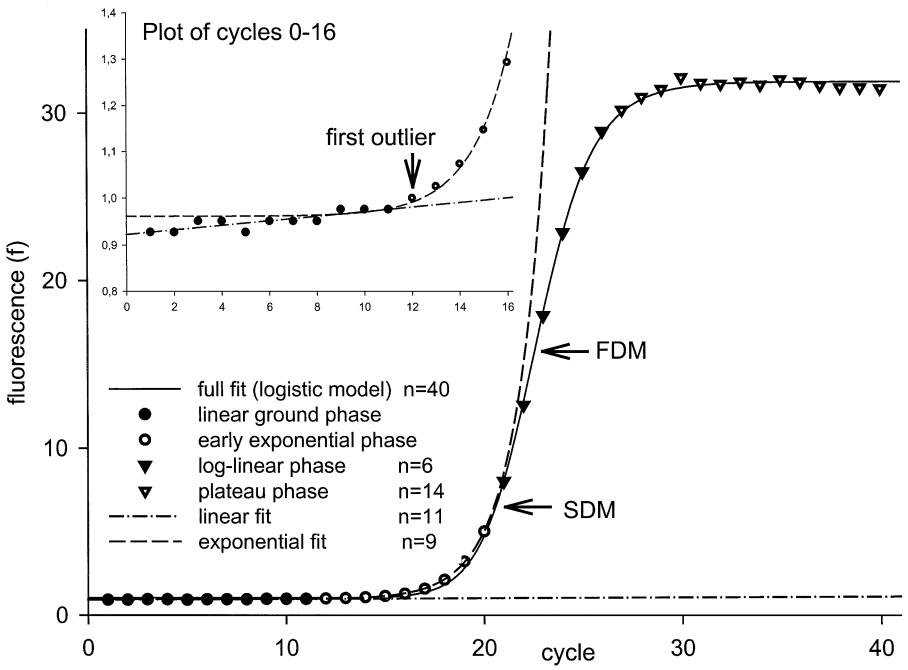


Figure 1. Plot of fluorescence observations from LightCycler (Roche Diagnostics). Forty observations give a sigmoid trajectory that can be described by a full data fit (FPLM). The ground phase can be linearly regressed (inlay). The following data of $n > 7$ are considered to behave exponentially and can be fitted using an exponential model. Various model fits are described in the legend within the figure. FDM and SDM denote the position of the FDM and SDM within the full data fit.

ground phase or not, standardized residuals of the linear regression are computed. The last one of the regressed observations is always inspected as to whether it does or does not deviate from the linear trend. This procedure starts with the first three observations and proceeds in the way shown in the flowchart in Figure 2.

Computation of the studentized residual statistics is a way to obtain a test on the distribution of particular residuals. To test statistically the probability that a given residual value is an outlier we must ensure that the residual value is comparable with some defined pre-existing probability distribution (here a t_{n-1-p} distribution; see later).

Since observation of x_i from the data set of varying n is always inspected, it must be taken into account that observations further from the $\bar{x}$ (mean value) have stronger influence [h_{ii} (leverage)] on the slope of the regression line:

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \mathbf{3}$$

Therefore, h_{ii} (equation **3**) is the measure of a particular influence of the respective observation x_i on the slope of the regression line.

Furthermore, the so called 'externally studentized' residual (24) is computed as follows:

$$r_{i(n-1)} = \frac{\varepsilon_i}{s_{(n-1)}\sqrt{1 - h_{ii}}} \qquad \mathbf{4}$$

where ε_i is the raw residual value, etc., the difference between the observed fluorescence (y_i) value and fitted fluorescence ($\hat{y}_i$)

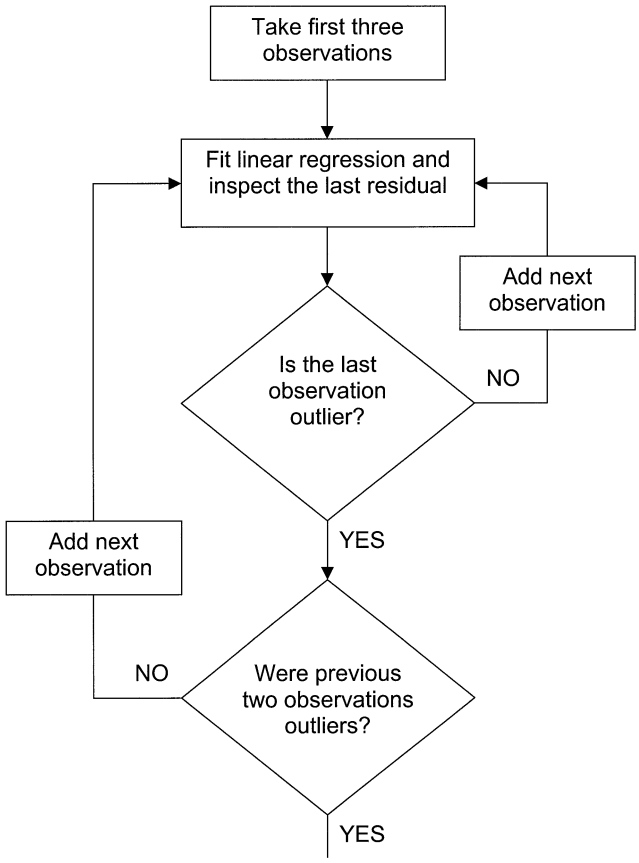


Figure 2. Flowchart of the statistical estimation of the beginning of the exponential phase based on inspection of externally studentized residuals.

value, $s_{i(n-1)}$ is the deviance of residuals in the regression model fitted over data with the deleted inspected observation $(n-1)$. This is computed as follows:

$$s_{i(n-1)} = \sqrt{MSE_{i(n-1)}} = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n-2}} \quad \mathbf{5}$$

$MSE_{i(n-1)}$ (equation **5**) is the mean square residual of the regression model with the deleted inspected data point. $n-2$ in the denominator denotes the residual degrees of freedom of the regression model.

Each $r_{i(n-1)}$ is distributed as t_{n-1-p} under the model. Therefore, we can test the hypothesis whether a single observation deviates from the model by comparing $r_{i(n-1)}$ with the t_{n-1-p} distribution (equation **6**) where $F(\cdot)$ is the cumulative distribution function of the t_{n-1-p} distribution:

$$P\text{-value} = 2 \times [1 - F(1 - |r_{i(n-1)}|)] \quad \mathbf{6}$$

Note, that even if the model holds for every observation (i.e. there are no outliers), one expects ~5% of the observations to have P-values <0.05. Therefore, we cannot automatically call every observation with a $P < 0.05$ an outlier, especially when n is large. If the observation is really an outlier and the fluorescence data points are entering the exponential phase, the following observations will also be detected as residuals. Based on experience, two more data points should be inspected after the first outlier is indicated to make sure that a consistent trend takes place (Fig. 1).

Determination of exponential observations

The start of the exponential behavior of the kinetic PCR is estimated by the described 'externally studentized' residual algorithm. We considered the end of the exponentially behaving observation to be just under the second derivative maximum (SDM) value as generated by LightCycler software 3.3 (Roche Diagnostics). Alternatively, from a four parametric logistic model (FPLM) with the parameters y_0, a, x_0 and b (equation **7**), fitting all fluorescence observations without any background correction gives:

$$f(x) = y_0 + \frac{a}{1 + \left(\frac{x}{x_0}\right)^b} \quad \mathbf{7}$$

where f is the value of function computed (fluorescence at cycles x), y_0 is the ground fluorescence, a is the difference between the maximal fluorescence acquired in the run and the ground fluorescence, x is the actual cycle number, x_0 is the first derivative maximum (FDM) of the function or the inflexion point of the curve and b describes the slope of the curve at x_0. The FPLM maximal value of its second derivative (SDM) is computed as follows. First, second and third derivatives of the model are calculated (data not shown). To result in an SDM, the third derivative must be null, which can be achieved by computing equation **8**. Two maxima are obtained; only the first 'positive maximum' is relevant for the approximation of the CP:

$$x_{SDM} = x_0 \cdot^b \sqrt{\frac{2 \cdot (1 - b^2) - \sqrt{3b^2 \cdot (b^2 - 1)}}{-b^2 - 3b - 2}} \quad \mathbf{8}$$

Other ways of computing the SDM were tested: these were based on just a distinct part of amplification trajectory around the expected SDM (25) or on a four parametric sigmoidal model (FPSM). These methods yield similar results to the SDM of FPLM values obtained (26) herein.

Estimation of amplification efficiency (E)

Once the beginning and the end of the exponential phase are defined, the exponential model is fitted over these data (equation **9**):

$$f = \gamma_0 + \alpha E^n \quad \mathbf{9}$$

The fluorescence value is represented by f, γ_0 is the upward shift due to ground fluorescence, α is the fluorescence due to the nucleic acid input, n is the cycle number and E is the efficiency of amplification in the early exponential phase of real-time PCR.

Verification of the method

Real-time PCR amplification efficiency was calculated from the given slopes in LightCycler Software 3.3 (Roche Diagnostics) (11). In the DNA calibration curve model, the efficiency per cycle was $E1_{fp} = 1.95$, using the 'fit-point method' (Table 1). The threshold fluorescence Y of the amplified real-time PCR product was calculated according to equation **10**:

$$Y = I * E^{CP} \quad \mathbf{10}$$

This resulted in a distinct product threshold fluorescence Y at a mean concentration ($n = 18$) of 9.91×10^{10} ss SRY molecules/set-up for $E1_{fp}$, with a coefficient of variance (CV) of Y of 79.65%. Additionally, the SDM in the LightCycler Software 3.3 (Roche Diagnostics) (11) was performed, and resulted in 2.89×10^{11} ss SRY molecules/set-up for $E1_{SDM}$ and in lower real-time PCR efficiency ($E1_{SDM} = 1.92$) and variation (CV = 41.48%).

Furthermore, the method of absolute fluorescence increase in the FDM (or point of inflection) of the amplification trajectory $E2_{FDM}$ (22) and in its SDM $E2_{SDM}$ was applied to compute the amplification efficiency E2. Briefly, the slope (or the first derivative) of the model curve at the respective maximum point is divided by the absolute fluorescence value reached at this point. These efficiencies varied between $1.351 < E2_{FDM} < 1.377$ and $1.448 < E2_{SDM} < 1.484$, with CVs of 159.77 and 195.92%, respectively. Y was also calculated and resulted in significant lower concentrations of 5.93×10^8 ss SRY molecules/set-up for $E2_{FDM}$ and 1.99×10^8 ss SRY molecules/set-up for $E2_{SDM}$. The difference between the general efficiency calculation methods E1 and E2 is approximately three orders of magnitude.

Finally, in the new single curve estimation method by FPLM, as suggested here, the mean product threshold fluorescence was 1.05×10^{11} ss SRY molecules/set-up with a variation of 30.80%, comparable with E1 methods. The calculated efficiency values varied in the range $1.822 < E_{new} < 1.884$, and lay between the evaluations described previously.

Table 1. Comparison of five different methods for the calculation of real-time PCR efficiencies

Conc.	n	CP_{fp}	CP_{SDM}	$E1_{fp}$ E_{all}	Y	CV% (Y)	$E1_{SDM}$ E_{all}	Y	CV% (Y)	$E2_{FDM}$ E	CV% (E)	Y	CV% (Y)	$E2_{SDM}$ E	CV% (E)	Y	CV% (Y)	E_{new} E	CV% (E)	Y	CV% (Y)
2.65E+07	3	11.02	14.10		8.58E+10	138.40		2.67E+11	0.23	2.59E+09	5.49	1.47	0.19	1.04E+09	1.47	1.84	0.40	1.43E+11	7.46
2.65E+06	3	15.93	17.20		1.10E+11	28.62		2.03E+11	0.16	6.74E+08	1.99	1.47	0.17	1.35E+08	0.42	1.85	0.67	1.04E+11	11.96
2.65E+05	3	18.47	20.53		5.82E+10	16.70		1.79E+11	0.22	2.02E+08	7.92	1.48	0.25	1.72E+07	1.59	1.85	0.28	7.88E+10	5.64
2.65E+04	3	21.45	24.88		4.24E+10	15.15		3.09E+11	0.37	7.25E+07	7.52	1.47	0.14	2.20E+06	1.33	1.86	1.59	1.36E+11	30.54
2.65E+03	3	26.08	28.18		1.25E+11	69.40		2.67E+11	14.56	1.36	0.48	1.83E+07	7.45	1.46	0.81	2.55E+05	1.21	1.84	1.34	7.71E+10	24.79
2.65E+02	3	30.31	32.66		1.74E+11	65.65		5.09E+11	24.13	1.36	0.38	6.28E+06	7.91	1.46	0.58	3.04E+04	1.09	1.83	0.15	9.25E+10	24.72
Summary for n = 18				**1.95**	**9.91E+10**	**79.65**	**1.92**	**2.89E+11**	**41.48**	**1.37**	**0.46**	**5.93E+08**	**159.77**	**1.47**	**0.71**	**1.99E+08**	**195.92**	**1.84**	**0.62**	**1.05E+11**	**30.80**

Conc., input concentration of nucleic acid in sample; CP_{fp}, CP based on the fit-point method; CP_{SDM}, CP based on the SDM computing method by LightCycler software 3.3 (Roche Diagnostics); $E1_{fp}$, amplification efficiency computed from calibration curve (11) where CPs are obtained as fit-points; $E1_{SDM}$, amplification efficiency computed from calibration curve where CPs are computed as the SDM; $E2_{FDM}$, amplification efficiency computed from absolute fluorescence increment in point of inflexion (FDM) of amplification trajectory (22); $E2_{SDM}$, amplification efficiency computed from absolute fluorescence increment in the SDM of amplification trajectory model; E_{new}, amplification efficiency computed according to the method suggested here; E, the mean value(s) of efficiency for n = 3; Y, fluorescence product computed from equation **10** for respective E for n = 3; CV, coefficient of variation for n = 3; Summary, either the overall mean or overall CV for n = 18.

The verification method was straightforward and was based on equation **10**. At the same threshold level, the amount of nucleic acids must also be identical in samples with a different known input concentration of nucleic acid. Here, the fractional value of n is known as the CP. If equation **10** is computed for each sample with the respective value of I, n and E, identical P-values should be theoretically obtained. The Y values were computed for each three concentrations of the dilution series used. As the E values obtained from different computing methods were entered, the method with the lowest variance of computed Y was considered the most accurate (Table 1).

DISCUSSION

As shown in Figure 1, fluorescence observations acquired from real-time PCR fluorescence monitoring are generally of a logistic or sigmoid shape (21,26), indicating that the PCR kinetics (27) consist of early ground phase, exponential growth phase, linear growth phase, and plateau phase. In the ground phase, the fluorescence acquisition is not detectable or just barely detectable due to the fluorescence passively emitted by the initial reaction system itself. At a certain cycle, the fluorescence emitted by the reaction product steps over the ground phase and enters the phase of growth. This phase takes several cycles and possesses a non-linear character (11). At the very beginning of this phase, the nature of the product increment can be well approximated as exponential ($r > 0.999$, $P < 0.001$). The rate of product generation slows down until the plateau phase is entered. In this phase, no more significant specific product is generated, as a consequence of reaction exhaustion (28).

Herein, we propose a method of real-time PCR amplification efficiency estimation based on single reaction kinetic observations. As shown in the theoretical work of Peccoud and Jacob (29), if the raw fluorescence observations on the PCR trajectory are available, they contain information about the amplification efficiency in itself. The pitfall in such an amplification efficiency estimation from fluorescence observations is that just a few of the reaction observations represent the initial exponential mode of the reaction. To detect where the reaction leaves its undetectable ground phase, a statistical method of residual inspections was applied. This method was robust enough to detect the first observation significantly diverging from the ground phase. In this method, no influence of the number of observations (n) was present, as long as very few observations were not inspected (n = 4). Such reaction kinetics, where the exponential phase is entered after the first three cycles, are, however, far from real usage.

The end of the exponentially behaved observations was placed into the last observation just before the SDM. This is not an arbitrary decision. After the reaction reaches the SDM, it weakens and looses its exponential character. The computing of the fractional value of the SDM for the purpose of efficiency estimation need not be of the precision demanded for threshold placement, because just discrete observations are used for the efficiency computing. In this respect, the fit of the full-observation model such as the FPLM can be used for computing the SDM. Taking LightCycler computed values of SDM (23) yields similar results. Such a delimitation of observations representing the exponentially behaved part of the PCR yields a set of observations that can be fitted by an

exponential model (equation **10**) with high significance ($r^2 > 0.999$), where the number of raw data fluorescence observations per set-up was $n > 7$.

This efficiency calculation method was tested on a further four bovine target sequences of IGF-1, TNFα, prion protein and 18S rRNA amplified in several independent runs on the LightCycler platform (Roche Diagnostics), and resulted in similar findings (data not shown). Furthermore, the method was applied to real-time fluorescence data generated during the amplification of the recombinant sequence of the *Pyrenophora teres* 18S rRNA gene on an ABI-Prism 7700 instrument (Applied Biosystems, Branchburg, USA), using either SYBR Green I dye or a FIC-labeled minor groove binding 18S rRNA probe. Good results comparable with the dilution series method were also obtained here (data not shown). Altogether, 145 reaction kinetics of various samples have been analyzed in this way, all giving consistent results.

This shows that the presented algorithm is independent of the used platform, the used fluorescence dye (SYBR Green I or FIC), the analyzed target gene and, furthermore, independent of any arbitrary decisions made by the investigator.

In conclusion, verification recalculation of the product amount at a constant threshold level of fluorescence with known efficiency showed that such computed efficiency is more accurate than the method currently used. This is above all clear in the dilution series of the same sequence as the method shows the resolution for various input target concentrations in the sample. Such a computed amplification efficiency can be output from automated platforms, as well as CP values for each sample. This is the major advantage in contrast to other real-time PCR efficiency calculation methods (11,21,22). Efficiency estimations done after the SDM are therefore underestimating the real-time PCR efficiency, whereas previously described methods using a dilution series overestimate it. The newly developed method, with values lying between those of the conventional methods, in our opinion, reflects the 'real PCR efficiency'. The CV values for the variation of Y might seem to be too large (e.g. CV for $E1_{fp}$ = 79.65%; Table 1). Here, the fact must be taken into account that a great deal of the Y variance is caused by initial vertical shifts in the ground phase. That is, different samples have different fluorescence products already at the very beginning, before any cycling starts. This discrepancy between different samples contributes to the overall CV value for a given method (Table 1). Therefore, not the absolute CV values, but rather its order, is a measure of the applicability of a given method.

Although we want to stress the possibility of determining the amplification efficiency from just a single sample, a statistical approach with more replicates can be adopted. Herein, three replicates were investigated to confirm the stability of the described model.

REFERENCES

1. Liss,B. (2002) Improved quantitative real-time RT–PCR for expression profiling of individual cells. *Nucleic Acids Res.*, **30**, E89.
2. Higuchi,R., Fockler,C., Dollinger,G. and Watson,R. (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology*, **11**, 1026–1030.
3. de Silva,D. and Wittwer,C.T. (2000) Monitoring hybridization during polymerase chain reaction. *J. Chromatogr. B Biomed. Sci. Appl.*, **741**, 3–13.
4. Holland,P.M., Abramson,R.D., Watson,R. and Gelfand,D.H. (1991) Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl Acad. Sci. USA*, **88**, 7276–7280.
5. Whitcombe,D., Theaker,J., Guy,S.P., Brown,T. and Little,S. (1999) Detection of PCR products using self-probing amplicons and flourescence. *Nature*, **17**, 804–807.
6. Morrison,T.B., Weis,J.J. and Wittwer,C.T. (1998) Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques*, **24**, 954–958.
7. Marras,S.A.E., Kramer,F.R. and Tyagi,S. (1999) Multiplex detection of single-nucleotide variations using molecular beacons. *Genet. Anal.*, **14**, 151–156.
8. Schmittgen,T.D. (2001) Real-time quantitative PCR. *Methods*, **25**, 383–385.
9. Klein,D. (2002) Quantification using real-time PCR technology: applications and limitations. *Trends Mol. Med.*, **8**, 257–260.
10. Meuer,S., Wittwer,C. and Nakagawara,K. (2001) *Rapid Cycle Real-time PCR: Methods and Applications*. Springer, Heidelberg.
11. Rasmussen,R. (2001) Quantification on the LightCycler instrument. In Meuer,S., Wittwer,C. and Nakagawara,K. (eds), *Rapid Cycle Real-time PCR: Methods and Applications*. Springer, Heidelberg, pp. 21–34.
12. Thellin,O., Zorzi,W., Lakaye,B., De Borman,B., Coumans,B., Hennen,G., Grisar,T., Igout,A. and Heinen,E. (1999) Housekeeping genes as internal standards: use and limits. *J. Biotechnol.*, **75**, 291–295.
13. Warrington,J.A., Nair,A., Mahadevappa,M. and Tsygantskaya,M. (2000) Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol. Genomics*, **2**, 143–147.
14. Schmittgen,T.D. and Zakrajsek,B.A. (2000) Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT–PCR. *J. Biochem. Biophys. Methods*, **46**, 69–81.
15. Suzuki,T., Higgins,P.J. and Crawford,D.R. (2000) Control selection for RNA quantitation. *Biotechniques*, **29**, 332–337.
16. Pfaffl,M.W. (2001) A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res.*, **29**, E45.
17. Pfaffl,M.W., Horgan,G.W. and Dempfle,L. (2002) Relative Expression Software Tool (REST©) for group wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.*, **30**, E36.
18. Meijerink,J., Mandigers,C., van de Locht,L., Tonnissen,E., Goodsaid,F. and Raemaekers,J. (2001) A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR. *J. Mol. Diagn.*, **3**, 55–61.
19. Wittwer,C.T., Ririe,K.M., Andrew,R.V., David,D.A., Gundry,R.A. and Balis,U.J. (1997) The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques*, **22**, 176–181.
20. Livak,K.J. (2001) ABI Prism 7700 Sequence Detection System User Bulletin #2. Relative quantification of gene expression. http://docs.appliedbiosystems.com/pebiodocs/04303859.pdf
21. Liu,W. and Saint,D.A. (2002) A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Anal. Biochem.*, **302**, 52–59.
22. Pfaffl,M.W. (2001) Development and validation of externally standardised quantitative insulin like growth factor-1 (IGF-1) RT–PCR using LightCycler SYBR® Green I technology. In Meuer,S., Wittwer,C. and Nakagawara,K. (eds), *Rapid Cycle Real-time PCR: Methods and Applications*. Springer, Heidelberg, pp. 21–34.
23. Ririe,K.M., Rasmussen,R. and Wittwer,C.T. (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal. Biochem.*, **245**, 154–160.
24. Neter,J., Kutner,M.H., Nachtsheim,C.J. and Wasserman,W. (1996) *Applied Linear Statistical Models*, 4th Edn. Irwin, Chicago.
25. Wittwer,C.T., Gutekunst,M. and Lohmann,S. (1999) Method for quantification of an analyte. United States Patent No. US 6,303,305 B1.
26. Tichopad,A., Dzidic,A. and Pfaffl,M.W. (2002) Improving quantitative real-time RT–PCR reproducibility by boosting primer-linked amplification efficiency. *Biotechnol. Lett.*, **24**, 2053–2057.
27. Schnell,S. and Mendoza,C. (1997) Theoretical description of the polymerase chain reaction. *J. Theor. Biol.*, **188**, 313–318.
28. Kainz,P. (2000) The PCR plateau phase—towards an understanding of its limitations. *Biochim. Biophys. Acta*, **1494**, 23–27.
29. Peccoud,J. and Jacob,C. (1998) Statistical estimation of PCR amplification rates. In Ferré,F. (ed.), *Gene Quantification*. Birkhauser, Boston, pp. 111–128.

6688 *Nucleic Acids Research, 2003, Vol. 31, No. 22*
DOI: 10.1093/nar/gkg885

ERRATUM

Standardized determination of real-time PCR efficiency from a single reaction set-up

Ales Tichopad, Michael Dilger, Gerhard Schwarz and Michael W. Pfaffl

Nucleic Acids Research (2003) **31,** e122

The publishers would like to apologise for Figure 2 being incomplete. The complete and correct figure appears below.

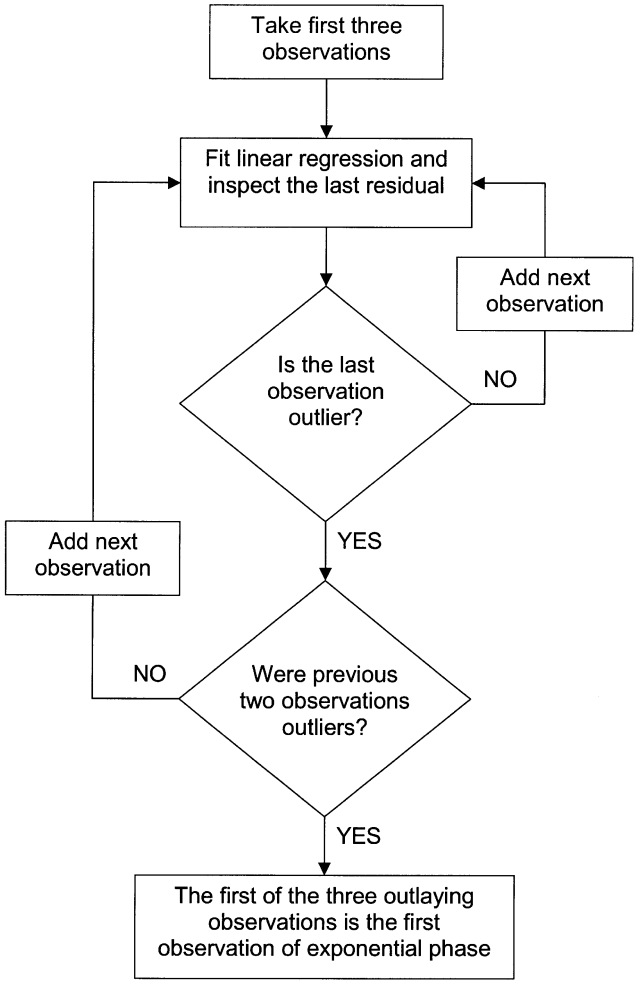


Figure 2. Flowchart of the statistical estimation of the beginning of the exponential phase based on inspection of externally studentized residuals.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular and Cellular Probes 17 (2003) 5–10
MOLECULAR AND CELLULAR PROBES
www.elsevier.com/locate/ymcpr

Tissue-specific expression pattern of bovine prion gene: quantification using real-time RT-PCR

Ales Tichopad, Michael W. Pfaffl, Andrea Didier*

Institute of Physiology, FML Weihenstephan, Technical University of Munich, Weihenstephaner Berg 3, 85354 Freising, Germany

Received 8 May 2002; revised 4 November 2002

Abstract

In recent studies PrP mRNA was determined mostly by in situ hybridisation or Northern Blot analysis—methods not suitable for absolute quantification of mRNA copy numbers. Herein we report on bovine prion mRNA quantification using calibrated highly sensitive externally standardized real-time RT-PCR with LightCycler instrument. Total RNA was isolated from nine different regions of the CNS and seven peripheral organs. PrP^c mRNA copy numbers could be determined in all tissues under study. In approval with prior studies high mRNA level was found in Neocortex and Cerebellum. Lymphatic organs showed at least as high expression levels of prion mRNA as overall brain. Lowest expression was detected in kidney. Results of our study provide insight into the involvement of different organs in pathogenesis with respect to prion mRNA expression. LightCycler technology is currently considered the most precise method for nucleic acid quantification and showed to be powerful tool for further studies on prion diseases pathogenesis.

Keywords: Prion; Absolute quantification; Cattle; mRNA expression; Real-time RT-PCR

1. Introduction

Cellular prion protein (PrP^c) [1] a glycosylphosphatidyl inositol (GPI) anchored glycoprotein [2] expressed in numerous cell types [3] and tissues is suspected to be involved in the pathogenesis of prion diseases [4,5]. These neurodegenerative disorders are described in many species such as cattle (BSE), sheep (Scrapie), mink (TME), cats, (FSE) and also in humans (CJD) (for overview see 5). Alterations are histopathologically characterised by accumulation of pathogenic prion protein (PrP^{sc}) isoform. During disease progression PrP^c serves as a substrate molecule for PrP^{sc} that acts as a template [4–6]. Due to direct interaction between these two types of molecules PrP^c undergoes autocatalytic conformational changes and turns PrP^{sc}. Unlike PrP^c that has a more α-helical content, PrP^{sc} mainly shows β-sheeted structure [7,8]. As no other pathogens or nucleic acids seem to be involved in this process, it is called the 'protein-only hypothesis' [9].

Pathological alterations are mostly related to the central nervous system (CNS), but some early studies indicated that in pre-clinical stages of disease progression peripheral organs might play a crucial role [10–12] in pathogenesis. In this regard lymphoid organs are already of long-term high interest [13–14].

Expression of prion gene in neuronal and non-neuronal tissues has to be taken into special consideration for a better understanding of its role in organism as well as in prion disease pathogenesis and for consumption risk assessment. Spread of PrP^{sc} from peripheral organs to the CNS is poorly understood to date. Nevertheless it becomes more apparent, that cells of the immune system play an important role in PrP^{sc} accumulation and distribution [15,16]. Amount of PrP^c mRNA in these cells and subsequent translation product abundance probably play a role in disease initiation and progression. It is likely that cells with higher expression of PrP^c pose higher risk of conversion to and accumulation of PrP^{sc}.

Real-time RT-PCR using SYBR Green I technology [17] provides an excellent and highly sensitive method for absolute quantification of mRNA expression. Using an external calibration curve based on plasmid DNA the quantification model showed higher sensitivity, exhibited a larger quantification range, had a higher reproducibility, than models using recombinant RNA or diluted PCR product as calibration curve [18].

* Corresponding author. Tel.: +49-8161-714202; fax: +49-8161-715380.
E-mail address: didier@wzw.tum.de (A. Didier).

0890-8508/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0890-8508(02)00114-7

Getting insight into prion gene expression in tissues and organs possibly under various treatments is an essential starting point for further study of protein conversion and PrP^{sc} accumulation. Tissue-specific expression pattern determined with high reproducibility and accuracy is also essential for understanding poorly explained natural role of prions in organism. Herein we show results concerning prion mRNA expression in CNS and peripheral organs as well as we test suitability of above-mentioned method.

2. Material and methods

Three healthy male Holstein-Frisian calves at the age of six month and three healthy adult 'Brown Swiss' cows were selected for tissue material sampling. Animals were slaughtered using ordinary procedure according to EU's established hygienic policy at the Bayerisch Landesanstalt für Tierzucht, Grub. Following organs were sampled: Neocortex, Cerebellum, Thalamus, Hypothalamus, Pituitary gland, Medulla oblongata, Pars cervicalis, Pars thoracalis and Pars lumbalis, this all with respect to CNS. Concerning peripheral organs, samples of bronchial lymph nodes, spleen and thymus with respect to lymphoid organs together with muscle, liver, kidney and lung were gathered. For each region samples from minimally three animals were taken (see Table 1), immediately frozen in liquid nitrogen and then stored in -80 °C until RNA extraction procedure was performed.

Total RNA was extracted with commercially available preparation peqGOLD TriFast (Peqlab; Germany) utilizing single step modified liquid separation procedure [19]. Constant amounts of 1000 ng of RNA were reverse-transcribed to cDNA using 200 units of MMLV Reverse Transcriptase (Promega; USA) according to the manufacturers instructions. Fifteen randomly chosen control samples without reverse-transcriptase (RT-negative) were assayed as negative controls for RT reaction.

For usage as a standard the prion gene was cloned into pCR®4-TOPO® vector using TOPO TA Cloning® Kit for Sequencing (Invitrogen; The Netherlands). This circular construct was linearized with NotI restriction endonuclease (MBI Fermentas, Lithuania) and its purity was inspected on a 1% agarose gel. For standard curve acquisition five serial dilutions of double stranded plasmid DNA ranging from 10^3 to -10^7 molecules were then prepared (2×10^3 to 2×10^7 plasmid DNA molecules).

To verify PrP containing plasmid, it was sequenced by MWG Biotech (Germany) and showed 100% homology to the sequence in EMBL and GenBank accession number AF117327.

All measuring of nucleic acid concentrations were done at OD_{260} nm on spectrophotometer (BioPhotometer® Eppendorf; Germany) with 220 – 1600 nm UVettes®. PrP primers flanking a 262 bp fragment were constructed as follows: Forward 5: AAC CAA GTG TAC TAC AGG CCA, Reverse 5: AAG AGA TGA GGA GGA TCA CAG. Conditions for PCR were optimized in a gradient cycler (Mastercycler Gradient; Eppendorf; Germany) and subsequently in LightCycler analyzing melting curve of product acquired. This was done with respect to primer annealing temperature, primer concentration, template concentration and number of cycles applied.

Real-time PCR using SYBR Green I technology [18] in LightCycler with the above-mentioned primers was carried out amplifying cDNA of biological sample, negative controls and five plasmid DNA standards. Master-mix was prepared as follows: 6.4 μl of water, 1.2 μl $MgCl_2$

Table 1
Parameters of quantification. *n*, number of samples. Where $n = 3$, only calf samples were available due to different slaughtering procedure; y(RNA), yield of RNA in 1 mg of tissue; CV, coefficient of variation; copy/RNA, number of PrP mRNA copies in 1 ng of total RNA; copy/tissue, number of PrP mRNA copies in 1 mg of tissue

Tissue	*n*	y(RNA) (ng)	CV%	Copy/RNA (molecules)	CV%	Copy/tissue (molecules)	CV%
Neocortex	6	617	10.9	66154	50.9	4.1×10^7	50.3
Cerebelum	6	735	22.9	40095	12.5	3.0×10^7	30.1
Thalamus	6	445	15.4	9408	50.9	4.2×10^6	51.5
Hypothalamus	6	388	29.3	5621	43.3	2.2×10^6	63.4
Pituitary gland	3	311	17.3	6379	68.2	2.0×10^6	74.1
Medulla oblong	3	336	15.5	17026	5.6	5.7×10^6	20.8
Pars cervicalis	3	298	17.3	20483	45.9	6.1×10^6	33.7
Pars thoracalis	3	388	51.7	12039	60.5	4.7×10^6	64.7
Pars lumbalis	3	308	43.2	11008	59.5	3.4×10^6	90.9
Spleen	3	2600	16.5	5911	8.8	1.5×10^7	24.6
Lymph nodes	3	3150	16.8	21360	16.9	6.7×10^7	25.9
Thymus	3	2580	41.4	4320	25.9	1.1×10^7	18.2
Muscle	6	352	31.3	5649	117.8	2.0×10^6	81.3
Liver	6	2990	14.5	3316	87.7	9.9×10^6	76.6
Kidney	6	1200	50.1	138	41.6	1.7×10^5	49.6
Lung	6	1320	30.7	1222	29.3	1.6×10^6	56

(25 mM), 0.2 μl of each primer (20 pmol), 1.0 μl Fast Start DNA Master SYBR Green I (Roche Diagnostics; Switzerland) mix. Nine μl of mastermix and 25 ng of reverse transcribed total RNA or plasmid DNA of the respective concentration. Following amplification program was applied: after 10 min of denaturation at 95 °C 40 cycles of 4-segment amplification were accomplished with: 15 s at 95 °C for denaturation, 10 s at 62 °C for annealing, 20 s at 72 °C for elongation and 5 s at 83 °C appended for a single fluorescence measurement above melting temperature of possible primer-dimers. This fourth segment eliminates a non-specific fluorescence signal and ensures accurate quantification of desired product. Subsequently, a melting step was performed consisting of 10 s at 95 °C, 10 s at 60 °C and slow heating with a rate of 0.1 °C per second up to 99 °C with continuous fluorescence measurement.

Quantification of PrP gene expression was performed in terms of PrP cDNA copies using LightCycler software 3.5 based on 'Second Derivative Maximum Method' (Roche Diagnostics; Switzerland). In this method a second derivative maximum within exponential phase of amplification curve is linearly related to a starting concentration of template cDNA molecules.

The mean, standard deviation (SD) and coefficient of variance (CV) was then calculated from obtained numbers of copies re-counted per 1 ng of total RNA and 1 mg of tissue for every organ and region over all six animals. Expression per mg of tissue was obtained as follows:

$$n_{\text{tissue}} = \text{y(RNA)} \times n_{\text{cDNA}}$$

where n_{tissue} is number of PrP^c mRNA copies in 1 mg of tissue, y(RNA) is for the yield of total RNA from 1 mg of tissue and n_{cDNA} means number of PrP cDNA copies in 1 ng of total cDNA. The distribution of all data sets was tested (Kolmogorov-Smirnov normality test) and, where necessary, data were normalized using common logarithm. Eventually one-way analysis of variance and *t*-tests were applied (SigmaStat; Jandel Scientific Software SPSS). Tissue-specific contrasts were inspected employing Tukey Test (SigmaStat; Jandel Scientific Software SPSS).

3. Results

Sensitivity of the LightCycler RT-PCR was evaluated using different starting amounts of mRNA and standard curve. SYBR Green I fluorescence determination at the elevated temperature 83 °C resulted in a reliable and sensitive cDNA quantification assay with high linearity (Pearson correlation coefficient = 0.99) over five orders of magnitude from 2×10^3 to 2×10^7 recombinant standard DNA start molecules (Fig. 1).

To verify real-time RT-PCR products derived either from plasmid or tissue total RNA a melting curve analysis on LightCycler (Roche) and gel electrophoresis were performed. Products showed no primer dimers, single sharp peak, identical melting points and expected length of 262 bp in gel electrophoresis.

Real-time PCR efficiencies were calculated from the given slopes (three repeats) in LightCycler Software 3.5

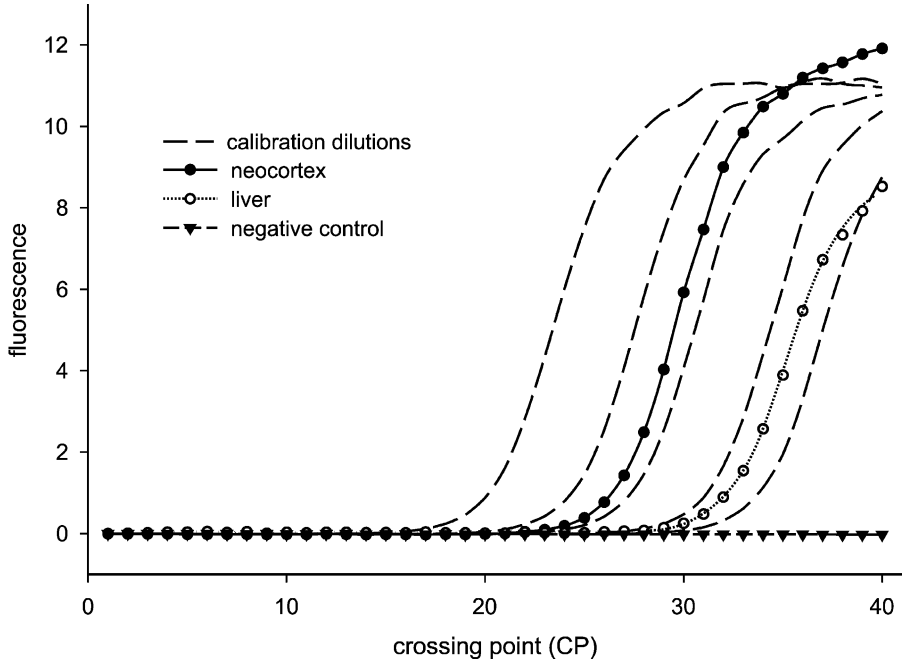


Fig. 1. Example of real-time PCR amplification curves obtained by plotting fluorescence data against their cycle number. Five calibration dilutions (2×10^7–2×10^3 copies) are shown together with two biological sample of Neocortex (4.25×10^5 copies) and liver (7.41×10^3 copies) and a negative control (without nucleic acid input).

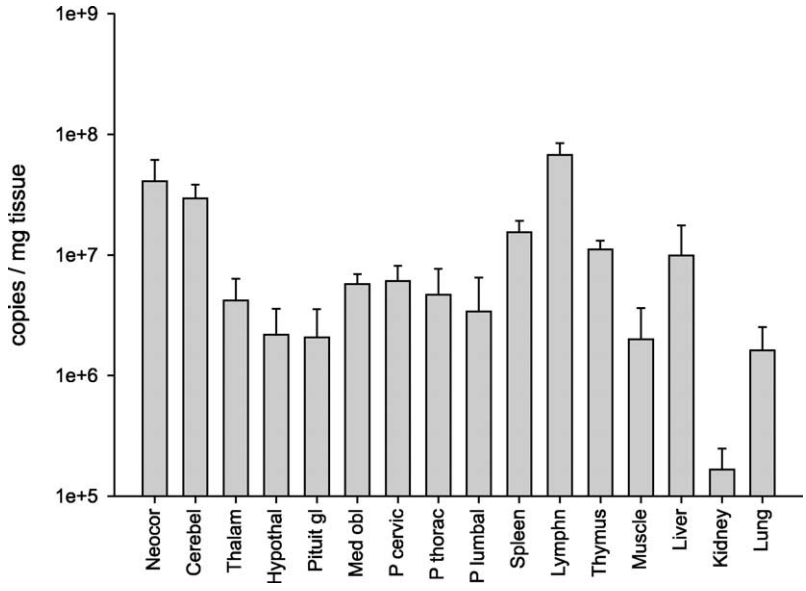


Fig. 2. Quantification of PrP mRNA copies in 1 mg of tissue. Error lines indicate SD.

(Roche) and showed efficiency rates per cycle in the DNA calibration curve model of $E = 1.90$. Sensitivity, linearity and reproducibility of the developed PrP real-time PCR assay allows for the absolute and accurate quantification of cDNA molecules even in tissues with low abundances down to a few molecules. After optimization of RNA extraction procedure no contamination by DNA was detectable in RT-negative samples.

Significantly ($p < 0.01$; log transformed) different amounts of total RNA were obtained ranging from 308 ng/mg (Pars lumbalis) to 3150 ng/mg of tissue (lymph nodes). For overview see Table 1. Inter-assay variation, between different real-time PCR runs, with the same cDNA ($n = 6$) was $<10\%$. Intra-assay variation, within one real-time PCR run, was $<5\%$. Inter-assay variation with the same total RNA but reverse-transcribed in different reactions had variance ranging from 7.2 to 30%.

No significant difference (t test, $p = 0.05$) in PrP RNA expression was found between both groups of animals concerning mean overall values and with respect to individual organs. For the purpose of statistical analysis the group of six animals was considered homogenous.

Significantly ($p < 0.01$; log transformed) different numbers of prion mRNA copies were determined in 1 ng of total RNA. The results of quantification are shown in Table 1. The highest PrP mRNA amount per 1 ng of total RNA was found in Neocortex. The lowest was found in kidney. As for the neuronal tissues, second highest levels of mRNA expression was found in Cerebellum. Lymph nodes showed highest content of PrP mRNA in 1 ng of total RNA within the three lymphoid organs studied. From other organs the highest copy number of PrP mRNA was determined in liver.

Re-counted to 1 mg of tissue, the highest expression level was found in lymph nodes and the lowest in kidney (Table 1, Fig. 2). Neocortex showed the highest expression within all CNS. Other neuronal tissues including different parts of spinal cord express similar levels of PrP. For tissue-specific contrast see Table 2.

4. Discussion and conclusions

Intuitive accord of our results with previously published data shows that real-time RT-PCR using LightCycler is reliable and powerful instrument to study pathogenesis of prion disease on mRNA level. Nevertheless, it makes these studies comparable and reproducible. It yields valued outputs as to levels of prion mRNA

Table 2
Tissue-specific contrasts in PrP mRNA copy number in 1 mg of tissue. The sign ‘ + ’ indicates significant ($p < 0.05$) difference, ‘ − ’ vice versa

p<0.05	Lymph n	Neocor	Cerebel	Spleen	Thymus	Liver	P cervic	Med obl	P thorac	Thalam	P lumbal	Hypothal	Pituit gl	Muscle	Lung	Kidney
Lymph n		-	-	-	-	+	+	+	+	+	+	+	+	+	+	+
Neocor	-		-	-	-	+	+	+	+	+	+	+	+	+	+	+
Cerebel	-	-		-	-	-	-	-	+	+	+	+	+	+	+	+
Spleen	-	-	-		-	-	-	-	-	-	-	+	+	+	+	+
Thymus	-	-	-	-		-	-	-	-	-	-	-	-	+	+	+
Liver	+	+	-	-	-		-	-	-	-	-	-	-	-	-	+
P cervic	+	+	-	-	-	-		-	-	-	-	-	-	-	-	+
Med obl	+	+	-	-	-	-	-		-	-	-	-	-	-	-	+
P thorac	+	+	+	-	-	-	-	-		-	-	-	-	-	-	+
Thalam	+	+	+	-	-	-	-	-	-		-	-	-	-	-	+
P lumbal	+	+	+	-	-	-	-	-	-	-		-	-	-	-	+
Hypothal	+	+	+	+	-	-	-	-	-	-	-		-	-	-	+
Pituit gl	+	+	+	+	-	-	-	-	-	-	-	-		-	-	+
Muscle	+	+	+	+	+	-	-	-	-	-	-	-	-		-	+
Lung	+	+	+	+	+	-	-	-	-	-	-	-	-	-		+
Kidney	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	

in different tissues. The reverse transcription reaction is the most significant cause of error. From our and other's experiences. We consider its efficiency between 30–70%, depending on sample type. Apparently, there is a general tissue-specific effect on quantification results. The RNA extraction procedure is also an important factor influencing quantification accuracy, if the copy number is expressed per tissue weight. This becomes clear after comparison of variance in PrP mRNA amounts in total extracted RNA and in tissue (Table 1). Concerning different RNA copy numbers per ng total RNA and per mg tissue, discussion has to face two different interpretations. Firstly, copy numbers per ng total RNA give good insight into total PrP^{c} expression potential of different organs. These data are more important for understanding of an early pathogenesis in prion disease. Secondly, the pattern of expression per tissue weight should have higher impact on consumption risk assessment of different organs. As mentioned in the introduction part, it is likely that organs with higher expression of PrP^{c} pose higher risk for conversion to and accumulation of PrP^{sc}. As daily food intake will be on gram level, PrP^{c} expression per mg tissue is important for consumer risk appraisal.

All three lymphatic organs express highly PrP mRNA, but fact that high total RNA yield was obtained from these tissues must be considered. This could possibly affect the re-calculation of copy number per weight of tissue. Nevertheless, our findings are consistent with role of PrP^{c} within immune system as suggested by Cashman [16] and with close prion-immune system linkage in general [15]. Nevertheless, they are in contrast to the earlier work of Robakis [20] where PrP was undetectable in normal rodent spleen. It is therefore necessary to focus more in detail on absolute PrP mRNA quantification in the above-mentioned organs and cells.

High PrP expression in neuronal tissues is consistent with works of Harris et al. [21,22] who were able to detect chicken prion protein mRNA in brain and a variety of organs by the means of in situ hybridisation and Northern Blot. But it has to be mentioned that above cited methods are semi-quantitative and under the detection abilities of real-time RT-PCR. Furthermore, they are not suitable for absolute quantification of PrP mRNA transcripts. Higher levels of expression in Neocortex and Cerebellum are coherent with several well-postulated hypothesis on PrP^{c} distribution in vertebrate organism and its ultimate role for normal neuronal function in CNS [23,24]. No apparent contrast in PrP mRNA levels was observed between white and gray matter of Neocortex, Medulla oblongata and Medulla spinalis (data not shown), although possible contamination during region segregation must be taken into account. In contrast, other regions of the brain as well as spinal cord showed no significant difference to other peripheral organs such as liver, muscle or lung. Kidney with its lowest expression was significantly different from all other tissues. Cells possibly responsible for higher expression levels in liver compared to muscle could be Kupffer's cells belonging to the antigen presenting subpopulation of white blood cells. Muscle and liver had different expression levels with more pronounced mRNA amount in the liver, which should have impact on consumption risk assessment. This gives a good intuitive sense, concerning that no PrP^{sc} infectivity has ever been detected within muscle tissue. We detected considerably low mRNA amounts in kidney compared to all above tissues. As we detected PrP mRNA in all tissues under study, it is probable that a post-transcriptional regulation finally determines PrP^{c} amount on protein level and its anchorage on the cell surface.

Acknowledgements

We thank the Bayerische Landesanstalt für Tierzucht at Grub for excellent technical support during the slaughtering and tissue sampling process.

References

[1] Prusiner SB. Novel proteinaceous infectious particles cause scrapie. Science 1982;216:136–44.
[2] Stahl N, Borchelt DR, Hsaio K, Prusiner SB. Scrapie prion protein contains a phosphatidylinositol glycolipid. Cell 1987;51:229–40.
[3] Brown HR, Goller NL, Rudelli RD, et al. The mRNA encoding the scrapie agent protein is present in a variety of non-neuronal cells. Acta Neuropathol 1990;80:1–6.
[4] Prusiner SB. Prions. Proc Natl Acad Sci USA 1998;95:13363–83.
[5] Liemann S, Glockshuber R. Transmissible spongiform encephalopathies. Biochem Biophys Res Commun 1998;250:187–90.
[6] Brandner S, Isenmann S, Raeber A, et al. Normal host prion protein necessary for scrapie-induced neurotoxicity. Nature 1996;379:339–43.
[7] Pan KM, Baldwin M, Nguyen J, et al. Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. Proc Natl Acad Sci USA 1993;90:10962–6.
[8] Safar J, Roller PR, Gajdusek DC, Gibbs CJ. Conformational transitions, dissociation, and unfolding of scrapie amyloid (prion) protein. J Biol Chem 1993;268:20276–84.
[9] Griffith JS. Self-replication and scrapie. Nature 1967;215:1043–4.
[10] Kimberlin RH, Walker CA. Pathogenesis of mouse scrapie: dynamics of agent replication in spleen, spinal cord and brain after infection by different routes. J Comp Pathol 1979;89:551–62.
[11] Kimberlin RH, Walker CA. Pathogenesis of scrapie (strain 263K) in hamsters infected intracerebrally, intraperitoneally or intraocularly. J Comp Pathol 1986;67:255–63.
[12] Kimberlin RH, Walker CA. Incubation periods in six models of intraperitoneally injected scrapie depend mainly on the dynamics of agent replication within the nervous system and not the lymphoreticular system. J Gen Virol 1988;69:2953–60.
[13] Fraser H, Dickinson AG. Pathogenesis of scrapie in the mouse: the role of the spleen. Nature 1970;226:462–3.
[14] Fraser H, Dickinson AG. Studies of the lymphoreticular system in the pathogenesis of scrapie: the role of spleen and thymus. J Comp Pathol 1978;88:563–73.
[15] Aucouturier P, Carp RI, Carnaud C, Wisniewski T. Prion diseases and the immune system. Clin Immunol 2000;96:79–85.

[16] Cashman NR, Loertscher R, Nalbantoglu J, et al. Cellular isoform of the scrapie agent protein participates in lymphocyte activation. Cell 1990;61:185–92.
[17] Morrison T, Weis JJ, Wittwer CT. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. BioTechniques 1998;24:954–62.
[18] Pfaffl MW, Hageleit M. Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT-PCR. Biotechnol Lett 2001;23:275–82.
[19] Chomczynski P. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. BioTechniques 1993;15:532–7.
[20] Robakis NK, Sawh PR, Wolfe GC, Rubenstein R, Carp RI, Innis MA. Isolation of a cDNA clone encoding the leader peptide of prion protein and expression of the homologous gene in various tissues. Proc Natl Acad Sci USA 1986;83:6377–81.
[21] Harris DA, Lele P, Snider WD. Localization of the mRNA for a chicken prion protein by in situ hybridization. Proc Natl Acad Sci USA 1993;90:4309–13.
[22] Harris DA, Falls DL, Johnson FA, Fischbach GD. A prion-like protein from chicken brain copurifies with an acetylcholine receptor-inducing activity. Proc Natl Acad Sci USA 1991;88:7664–8.
[23] Collinge J, Whittington MA, Sidle KC, et al. Prion protein is necessary for normal synaptic function. Nature 1994;370:295–7.
[24] Sakaguchi S, Katamine S, Nishids N, et al. Loss of cerebellar Purkinje cells in aged mice homozygous for a disrupted PrP gene. Nature 1986; 380:528–31.

Biotechnology Letters **24:** 2053–2056, 2002.

Improving quantitative real-time RT-PCR reproducibility by boosting primer-linked amplification efficiency

Ales Tichopad, Anamarija Dzidic & Michael W. Pfaffl*
Institute of Physiology, FML-Weihenstephan, Center of Life and Food Science, Technical University of Munich, Munich, Germany
**Author for correspondence (Fax: +49-8161-714204; E-mail: pfaffl@wzw.tum.de)*

Received 29 August 2002; Revisions requested 12 September 2002; Revisions received 7 October 2002; Accepted 8 October 2002

Key words: gene quantification, PCR efficiency, primers, RT-PCR

Abstract

The effect of primer selection on real-time polymerase chain reaction (RT-PCR) performance was tested. Primer sets of varying length of product were used to amplify the sequence of β-actin. Variability in length caused variability in RT-PCR performance. Kinetic parameters of PCR were studied by mathematical approximation of real-time data by means of a four-parametric sigmoid model. This model describes the full kinetics of the amplification trajectory. Statistical exploration of parameters yielded by this model revealed that reactions with higher amplification efficiency – primed by well-performing primers – proceed with lower variability and are therefore better suited for measurement purposes.

Introduction

Reverse transcription (RT) polymerase chain reaction (PCR) is, because of its sensitivity, the method of choice for quantifying low abundance mRNAs in cells and tissues (Schmittgen 2001, Gibson *et al.* 1996, Rasmussen 2001).

There are two major disadvantages of using PCR for measurements: first, since the initial information is amplified exponentially, any error is also amplified in the same way. Second, the inherent stochastic character of the chain reaction is responsible for some initial information loss during amplification and thus the reproducibility can never be 100% (Peccoud & Jacob 1996). Nevertheless, optimising PCR conditions and data processing can increase its reproducibility. A myriad of optimisation protocols have been published but not many of them mention the problem of reaction priming. Routinely, reaction conditions are considered good when the trajectory of reaction (fluorescence curve) is steep, although this does not imply that a worse performing but well standardized reaction (Pfaffl 2002, Meijerink *et al.* 2002) cannot yield good quantitative results. We have now tested if a higher amplification efficiency achieved by primer selection can improve the reproducibility during amplification and had therefore has a direct impact on the reliability of the assay.

Materials and methods

RNA extraction and RT-reaction

Tissue samples of liver, jejunum, heart and spleen from six sheep were stored in liquid N_2 after animals had been slaughtered. Total RNA was extracted with commercially available preparation TriPure (Roche Diagnostics). Constant amounts of 1000 ng RNA were reverse-transcribed to cDNA using 200 units Moloney Murine Leukemia Virus Reverse Transcriptase (Promega) according to the manufacturer's instructions.

PCR amplification

Sequences of β-actin gene coding region varying in length, summarized in Table 1, were amplified in

Table 1. Primer characteristics used to amplify five various sequences of β-actin gene coding region.

Primer[a]	Sequence	GC content (%)	Melting temperature (°C)	Product length (bp)
	for CAC GGA ACG TGG TTA CAG CTT TAC C	52	64	
P1				56
	rev TGT CAC GCA CAA TTT CCC GCT C	54	62	
	for GAA CGT GGT TAC AGC TTT AC	45	55	
P2				99
	rev ATC TCC TGC TCG AAG TCC A	53	57	
	for ATC CTC ACG GAA CGT GGT TAC AGC	54	64	
P3				159
	rev ATC GGG CAG CTC ATA GCT CTT CTC	54	64	
	for GTG CGT TGA CAT CAA GGA GAA GCT C	54	64	
P4				217
	rev TTG AAG GTG GTC TCG TGA ATG CCG	54	64	
	for AAG GCC AAC CGT GAG AAG ATG ACC	54	64	
P5				298
	rev TGT CAC GCA CAA TTT CCC GCT C	55	62	

[a]for = forward; rev = reverse.

25 ng cDNA in LightCycler instrument (Roche Diagnostics) (Wittwer *et al.* 1997). Five different beta-actin primer sets were used in all 24 samples (4 tissues $\times$ 6 animals) to generate variability in amplification kinetics. All primer sets were designed to generate PCR products from 50 to 300 bp in nearly equal steps of 50 bp difference. RT-PCR products should differ only in their length. The primer characteristics, like GC content and annealing temperature of all sets were adjusted to nearly constant values (Table 1).

The master-mix for each PCR run was prepared as follows: 6.4 μl water, 1.2 μl $MgCl_2$ (25 mM), 0.2 μl each primer (20 pmol), 1 μl Fast Start DNA Master SYBR Green I (Roche Diagnostics) mix, 9 μl of master-mix and 25 ng reverse transcribed total RNA. The following amplification protocol was used for all runs with all five primer sets: denaturation program (95 °C for 10 min), a three-segment amplification and quantification program repeated 40 times (95 °C for 15 s, 60 °C for 10 s and 72 °C for 20 s), melting curve program (95 °C for 10 s, 60 °C for 10 s and then slow heating with a rate of 0.1 °C per s to 99 °C with continuous fluorescence measurements) and finally cooling program down to 40 °C. All fluorescence measurements during the above mentioned program were performed in fluorescence acquisition channel 1 set at value 3. In order to prevent inter-assay variation, samples with the same primer set were always amplified within one run.

Data processing

Fluorescence observations of all samples were taken directly from LightCycler software 3 (Roche Diagnostics). Using SigmaPlot 2000 (SPSS Inc, Chicago, USA) they then were fitted with four-parametric sigmoid model

$$f = y_0 + \frac{a}{1 + e^{-\left(\frac{x - x_0}{b}\right)}}. \quad (1)$$

In Equation (1) f is the value of function computed (fluorescence at cycles x), y_0 is the ground fluorescence, a is the difference between maximal fluorescence acquired in the run and the ground fluorescence, e is the natural logarithm base, x is the actual cycle number, x_0 is the first derivative maximum of the function or the inflexion point of the curve and b describes the slope of curve (Figure 1). In this model, the smaller is the value of b, the higher is the amplification efficiency. All PCR kinetics data produced by this model fit were finally statistically processed in SigmaStat 2.0 (Jandel Corporation, San Rafael, USA).

Table 2. Influence of different primer sets and different tissues on amplification efficiency computed using two-way ANOVA.

Source of variance	DF[a]	SS[b]	MS[c]	F[d]	*P*[e]
Primer	4	0.354	0.0886	34.056	<0.001
Tissue	3	0.0338	0.0113	4.332	0.006
Interaction	12	0.0551	0.0046	1.764	0.065
Residual	1000	0.26	0.0026		
Total	119	0.703	0.0059		

[a]Degrees of freedom.
[b]Sum of squares.
[c]Mean square.
[d]F-test ratio.
[e]Probability.

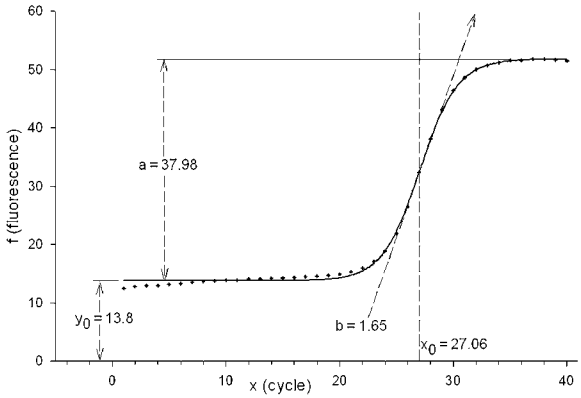


Fig. 1. Four-parametric sigmoid model. Model is described by Equation (1). One fluorescence data set from this study was used as an example. In this model, y_0 is the ground fluorescence, a is the difference between maximal fluorescence acquired in the run and the ground fluorescence, x_0 is the first derivative maximum of the function or the inflexion point of the curve and b describes the slope of curve.

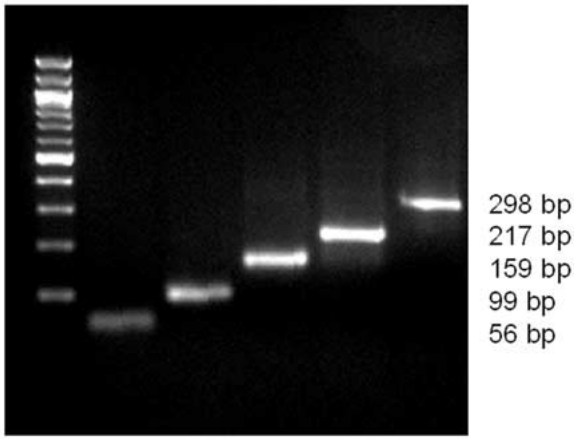


Fig. 2. Product purity inspection. cDNA sample from liver was amplified with five β-actin primer sets P1–P5 (lanes 1–5). Product lengths on 4% agarose gel match the theoretical designed length.

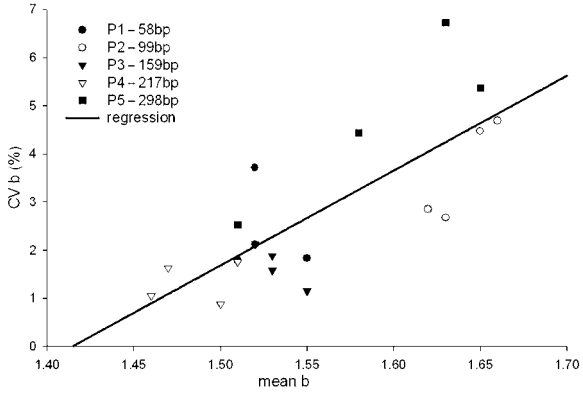


Fig. 3. Linear trend between mean value of parameter b and its variability. Each point represents six PCR runs with the same primer set and the same tissue-derived samples, but from different animals. Different primer sets are distinguished by various symbols. After fitting fluorescence data with a four-parametric sigmoid model, *mean b* (ordinate) and coefficient of *variance CV b* (abscissa) were computed and plotted.

Results and discussion

All five primer pairs generated highly specific products at the indicated lengths. Melting curve analysis (Ririe *et al.*1997) and gel inspection did not detect any primer dimers or other side-product (Figure 2). The applied mathematical model (Figure 1) is very suitable for so-called '*optimal PCR runs*' with high amplification efficiencies, low ground level fluorescence, high and constant plateau fluorescence, where the b value indicates very small variations (A. Tichopad, unpublished work). Its sensitivity is very much greater than other often used methods of serial dilutions or other methods of amplification efficiency detection (Liu & Saint 2002) As to our personal findings, the model gives justified fit and good b resolution if coefficient of determination $r^2 > 0.999$ (A. Tichopad, unpub-

lished work). Such a sigmoid estimator, however, must be considered relative as the slope of the curve does not directly indicate the amplification efficiency, e.g., the proportion between current and previous product amounts. Nevertheless, this point is not important if the various efficiencies are simply being compared and exact calculation of the differences is not required.

The significant trend ($P < 0.001$) shows that there is a lower variance of amplification efficiency in groups where b value is lower and therefore the PCR efficiency itself is higher. No such a trend is observed if variance of parameter a (e.g., high of plateau reached) was plotted against mean b. A two-way ANOVA test computed on b parameters of all runs with factors of *primer* and *tissue* showed that most of the variance between b parameters was caused by the primers (Table 2). There was also no trend between primer length and its amplification efficiency (Figure 3).

In LightCycler software, the second derivative maximum CP_{sdm} is an often used computing procedure to obtain the initial number of copies in sample. We can simulate this in the four-parametric sigmoid model, where second derivative maximum CP (SM_{sdm}) is computed from Equation (1) as follows: the first, the second, and the third derivations of the model are calculated (not shown). To calculate a second derivative maximum the third derivation has to be null: $f'''(x) = 0$. There are two second derivative maximums for $x \approx x_0 \pm 1,317 \cdot b$, whereas only the first '*positive maximum*' is relevant for an intelligent approximation of the CP. From the calculation, therefore:

$$CP(SM_{sdm}) \Rightarrow x = x_0 - 1,317 \cdot b. \qquad (2)$$

Equation (2) indicates a linear relationship between b and the value of the second derivative maximum CP (SM_{sdm}). Therefore, variability in CP (SM_{sdm}) is also linearly related to amplification efficiency in this simulation. The higher amplification efficiency, the lower variability of CP (SM_{sdm}).

To summarize and apply the above mentioned findings. Sample-specific factors such as fat, blood etc. as well as contaminants from extraction alter the PCR amplification parameters and hence introduce variability even when the same tissue samples are analysed but from different animals. We believe an error induced in this way can be minimized by boosting amplification efficiency. In this way, if compared samples undergoing PCR are forced towards their potential chemical-kinetic limits they can only vary over a smaller range. Therefore, when several primer sets are available, the set with highest amplification efficiencies should be chosen. It is likely that this concerns all other optimisation parameters (e.g., annealing temperature, Mg^{2+} concentration etc.). Several estimators of amplification efficiencies have been suggested. Here the suggested model is the most sensitive one for efficient reactions with steep trajectories. Also other sigmoid models can be used as relative estimators of amplification efficiency but their use must be considered and optimized according to the data analysed.

Acknowledgements

The authors thank F. Buckel for mathematical assistance. The experimental animals were slaughtered according to EU regulations at the EU official slaughterhouse: Bayerische Landesanstalt für Tierzucht at Grub, 85580 Poing, Germany.

References

Gibson UE, Heid CA, Williams PM (1996) A novel method for real time quantitative RT-PCR. *Genome Res.* **6**: 1095–1101.

Liu W, Saint DA (2002) A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Anal Biochem.* **302**: 52–59.

Meijerink J, Mandigers C, van de Locht L, Tonnissen E, Goodsaid F, Raemaekers J (2001) A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR. *J. Mol. Diagn.* **3**: 55–61.

Peccoud J, Jacob C (1996) Theoretical uncertainty of measurements using polymerase chain reaction. *Biophys. J.* **71**: 101–108.

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucl. Acids Res.* **29**: 2002–2007.

Rasmussen R (2001) Quantification on the LightCycler instrument. In: Meuer S, Wittwer C, Nakagawara K, eds. *Rapid Cycle Real-time PCR: Methods and Applications*. Heidelberg: Springer-Verlag Press, pp. 21–34.

Ririe KM, Rasmussen RT, Wittwer CT (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal. Biochem.* **245**: 154–160.

Schmittgen TD (2001) Real-time quantitative PCR. *Methods* **25**: 383–385.

Wittwer CT, Ririe KM, Andrew RV, David DA, Gundry RA, Balis UJ (1997) The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques* **22**: 176–181.

Submitted Manuscripts

Inhibition of *Taq* Polymerase and *MMLV* Reverse Transcriptase by Tea Polyphenols (+)-Catechin and (–)-Epigallocatechin-3- Gallate (EGCG)

Ales TICHOPAD[1], Jürgen POLSTER[2], Ladislav PECEN[3] & Michael W. PFAFFL[1*]

[1] Physiology Weihenstephan, Zentralinstitut für Ernährung- und Lebensmittel-forschung; [2] Department für 'Biowissenschaftliche Grundlagen', Lehrstuhl für Biologische Chemie, FG Physikalische Chemie, Technische Universität München, Wissenschaftszentrum Weihenstephan, D-85350, Freising, Germany; [3] First Medical Faculty of the Charles University,CZ-80200, Prague, the Czech Republic

**Author for correspondence:*
Michael W. Pfaffl
Tel: +49-8161-713511
Fax: +49-8161-714204
E-mail: pfaffl@wzw.tum.de

Abstract

Non-nutritional polyphenolic compounds such as (+)-catechin and (-)-epigallocatechin-3-Gallate (EGCG) are known as anticancer chemopreventive agents and have been utilised for medical purposes in form of tea drinking. Documented anticancer properties of these compounds result from their antioxidant effects. However, also direct alteration of an enzyme performance has been reported and deserves more attention. In this paper a direct effect of catechin and EGCG on the performance of reverse transcription (RT) and/or polymerase chain reaction (PCR) was studied. Both tea polyphenolic compounds were added into real-time RT-PCR reactions and the fluorescence data obtained were fitted with a mathematical model. Several parameters of PCR performance were compared obtained from the mathematical model for reactions with and without addition of (+)-catechin and EGCG. Addition of EGCG to enzyme reaction seems to inhibit the RT reaction ($p<0.05$) and to slow down the DNA polymerase reaction ($p<0.001$). Similarly, (+)-catechin inhibited the DNA amplification ($p<0.01$) but had no effect on the RT reaction. The effects could be observed in physiological flavanol concentrations ranging from 10^{-5} to 10^{-8} M.

Key words: Reverse Transcriptase; DNA Polymerase; Catechin; EGCG; real-time RT-PCR; Cancer.

Introduction

Polyphenolic compounds such as catechins or flavan-3-ols are supposed to be non-cytotoxic plant constituents for humans present in high concentrations in tea, but also in many other foods, such as apples, grapes, vine and their processed beverages. Beside the use as a beverage the medicinal properties of the green, black and Oolong tea were well known and utilised by many peoples in Asia. The long-term tea drinking habit of Asians is believed to be responsible for lowering cancer incidence. Recently, these compounds are under intensive scientific focus for their assumed anticancer preventive and curative properties (Lambert and Yang 2003, Middleton et al. 2000). However, evidence obtained from clinical trials and population studies, that would support such theories is conflicting. This is often a result of confounding factor correlated with tea drinking, such as smoking (Higdon and Frei 2003). As chemopreventive agents, these compounds are assumed to be able to halt or reverse the development and progression of tumour cells (Hayakawa et al. 2001, Morre et al 2003). It has been shown that EGCG can induce and also inhibit expression of genes associated with the cancer progression and apoptosis (Ahn et al. 2003, Gupta et al. 2002 and Okabe et al. 2001). Telomere shortening caused by this compound is discussed as well in literature (Seimiya et al. 2002) together with a possible interaction between histones and catechins (Polster et al. 2003). There is an abundant *in vitro* and less abundant *in vivo* evidence of a direct as well as indirect antioxidant effect of the tea polyphenols and their derivatives (Henning et al 2003, Cabrera et al. 2003, Wiseman et al. 1997). A direct effect such as the scavenging of reactive oxygen and nitrogen species is well studied. Inhibition of enzymes by tea polyphenols whose activity possibly increases oxidative stress in organism such as cytochromes (CYP); CYP P450 (Muto et al. 2001) and CYP 1A1 (Schwarz and Roots 2003) is an example of the indirect antioxidant effect.

The risk of oncogene changes in cells is linked to polymerase enzymes, therefore, a possible interaction between polymerases, their activities and catechin compounds deserves interest. Such an action was described earlier (Tao 1992), but this issue deserves surely more focus with an updated methodology and a greater public attention.

In addition, anti-viral effects of tea polyphenolic compounds are also discussed in recent literature (Fassina et al. 2002, Chang et al. 2003, Yamaguchi et al. 2002). Blocking of reverse transcriptase or just its partial inhibition may be at least a particular explanation of anti-retroviral curative properties of these compounds (Fassina et al. 2002). Preparations containing tea extracts have been already successfully clinically tested as effective *anti-Herpesvirus simplex* drug (commercial not published data).

Presumably, the mutation probability in metazoan organisms can be approximated by the polymerase chain reaction (PCR) with its exponential strengthening of the polymerase activity. However, using the updated RT-PCR systems, the analogy between the archeobacterial thermo-resistant DNA polymerase and the DNA polymerase in living metazoans is low. PCR performed on complementary DNA (cDNA) substrate, obtained by reverse transcription from the total RNA or mRNA, is a routinely used tool in expression analysis in a lot of laboratories today. Either for amplification of substrate aimed for later analysis, or as a direct precise analytical tool, PCR offers fast, easy and cheap way to quantify and classify selected sequences of nucleic acids. In the real-time PCR, not only final data on the quantity, but also data on the whole PCR kinetic performance is available (Wittwer et al. 1997). This additional information reports about the reaction system itself. Mathematical

approximation of this data by suitable model yields characteristic parameters of reaction kinetics. These parameters can be then compared between samples with experimental manipulation and control samples (Tichopad et al. 2002, Tichopad et al. 2003). Thus an attempt was started to simulate the reaction kinetics of the DNA polymerase (Wittwer et al. 1997) by real-time PCR on the LightCycler instrument (Roche Diagnostics, Basel, Switzerland). Additionally, the efficiency of the reverse transcriptase (RT) could be estimated. In this way, the influence of (+)-catechin and EGCG could be assessed on the *Thermus aquaticus (Taq)* DNA polymerase and on the *Moloney Murine Leukemia Virus (MMLV)* reverse transcriptase.

Materials and Methods

Tea polyphenols preparation

(+)-Catechin and (–)-epigallocatechin-3-Gallate (EGCG) were purchased from Sigma-Aldrich (Munich, Germany). Working dilution series in water of each compound to final concentrations 10^{-5}, 10^{-6}, 10^{-7}, and 10^{-8} M were prepared. Blood concentrations of 10^{-6} to 10^{-7} M represent physiological concentrations occurring in human blood (Middleton et al. 2000, Higdon and Frei 2003).

RT and PCR reaction preparation

Samples of bovine liver tissue were stored immediately after its sampling in liquid nitrogen and then in –80°C. Total RNA extraction was performed using the commercial preparation TriPure (Roche Diagnostics) utilising the single step extraction procedure according to Chomczynski (1993). The following two approaches of RT-PCR were adopted differing in their separation of the reverse transcription (RT) reaction from the polymerase chain reaction:

A) One-step real-time RT-PCR approach

In this approach, the studied flavanol compounds were added into reaction before starting the reverse transcription, and could thus influence it. The RT as well as the PCR were run together on the LightCycler platform (Roche Diagnostics) using the QuantiTect SYBR Green RT-PCR Kit (Qiagen, Hilden, Germany). Prior the PCR, the RT step was attached reverse transcribing 5 ng total RNA into cDNA. The reaction was set according to the standard protocol recommended by manufacturer (Qiagen). The master-mix was prepared as follows: 5 µl QuantiTect SYBR Green RT-PCR Master Mix, 0.5 µl forward primer (1 µM), 0.5 µl reverse primer (1 µM) and 0.1 µl QuantiTect RT Mix. 6.1 µl of master-mix was filled in glass capillaries and a 2.9 µl water containing 5 ng total RNA was added as RT-PCR template. Finally, always 1 µl water containing varying concentration of one of the two studied compounds was added into each capillary so that the final concentrations 10^{-5}, 10^{-6}, 10^{-7}, and 10^{-8} M were achieved. Primers were designed and purchased from MWG Biotech (Ebersberg, Germany) to amplify 359 bp long cDNA fragment of bovine TNFα (Tumour Necrosis Factor alpha), according to the literature (Wittmann et al. 2002). Capillaries with samples were closed, centrifuged and placed into a cycling rotor. A five-step experimental run protocol was used consisting of 1) RT step (20 min at 50°C), 2) denaturation step (15 min at 95°C); 3) amplification and quantification step repeated 45 times (15 s at 94°C; 10 s at 60°C; 20 s at 72°C; 5 s at 80°C with a single fluorescence measurement); 4) melting curve step (60°C to 99°C) with a heating rate of 0.1°C per s and a continuous measurement (Ririe et al. 1997); 5) fast cooling step down to 40°C.

Each concentration of one of the two studied flavanol compound was run in triplicate. In total 4 flavanol concentrations, 3 replicates, 2 compounds either (+)-catechin or EGCG (each n = 12) plus 8 control capillaries with reaction mix but without any flavanol were run in a single real-time RT-PCR run so that no

random inter-run effect was introduced into the experiment.

B) Two-step RT real-time-PCR approach

In this approach the mRNA was reverse transcribed with the *MMLV* reverse transcriptase into cDNA separately on another platform (TGradient, Biometra, Göttingen, Germany). The studied flavanols were added after the RT reaction. The two-step RT real-time PCR was employed additionally to study the effect of flavanol exclusively on the polymerase reaction without a possible effect on the RT. If the RT itself is affected, the performance of the following polymerase reaction would also be altered since the cDNA concentration at the start of the PCR differs between samples.

Within this approach the same, above-mentioned, 359 bp TNFα (Wittmann et al. 2002) was amplified using the LightCycler – FastStart DNA Master SYBR Green I (Roche Diagnostics). It is a ready-to-reaction mix for PCR that contains *Taq* DNA polymerase and DNA double-strand specific SYBR Green I dye for detection (Morrison et al. 1998). The polyphenols were added into PCR reaction diluted in varying concentration in 1 µl water. The concentrations of both compounds present in the PCR were the same like in the one-step approach. Triplicate design with 8 additional control reactions was adopted as well as in the one-step approach. All 32 samples were performed within one run.

Statistical analysis

Using SigmaPlot software (SPSS Inc., Illinois, USA) the four parametric sigmoid model (Tichopad et al. 2002), according to equation 1, was used to fit the fluorescence raw data observations (Figure 1) generated by LightCycler software version 3.5 (Roche Diagnostics).

$$f(x) = y_0 + \frac{a}{1 + e^{-\left(\frac{x - x_0}{b}\right)}} \qquad [1]$$

In total 64 reaction *kinetics curves* were obtained by fitting the fluorescence raw data of each single sample with the model. In the four-parametric sigmoid model, ***x*** is the cycle number, ***f(x)*** is the computed function of the fluorescence in dependence of cycle number ***x***, y_0 is the background fluorescence, ***a*** is the difference between maximal fluorescence reached at plateau phase and background fluorescence (i.e. the plateau height), ***e*** is the natural logarithm base, x_0 is the co-ordinate of the first derivative maximum of the model or inflexion point of the curve, and ***b*** describes the slope at x_0 in the log–linear phase.

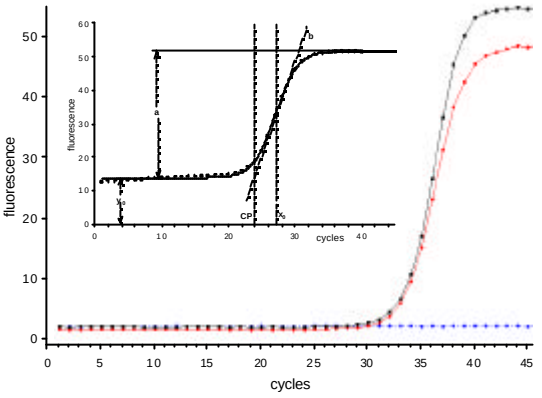


Figure 1. Real-time fluorescence history of two-step RT-PCR with flavanol treatment. (-■-) control group; (-●-) 10^{-6} M (+)-Chatechin treatment; (-●-) dotted baseline is no template water control.

Inlay: *Four-parametric sigmoid model (described by equation 1),* ***x*** *is the cycle number,* ***f(x)*** *is the computed function of the fluorescence at cycle* ***x***, y_0 *is the background fluorescence,* ***a*** *is the difference between maximal fluorescence reached at plateau phase and background fluorescence (i.e. the plateau height),* ***e*** *is the natural logarithm base,* x_0 *is the co-ordinate of the first derivative maximum of the model or inflexion point of the curve,* ***CP*** *is the co-ordinate of the second derivative maximum of the model, calculated by the LightCycler software 3.5 (Roche Diagnostics) and* ***b*** *describes the slope at* x_0 *in the log–linear phase.*

The following general scheme can be given:

a – high ***a*** detects high real-time PCR product obtained after all cycles,

b – low ***b*** detects high polymerase reaction efficiency (Tichopad et al. 2003),

x_0– high x_0 detects delay in the detectable phase of the reaction, comparable to ***CP***,
y_0 – high y_0 detects high RT product obtained.

Further parameter analysed was the crossing point (***CP***), which is also named ***Ct*** value. The parameter ***CP*** is not a part of the four-parametric sigmoid model but can be obtained by differentiation of the corresponding equation (1). It is a central term of the real-time PCR quantification (Wittwer et al.;1997; Rasmussen 2001). In brief, ***CP*** gives the number of PCR cycles at which the fluorescence generated by PCR product reaches a defined fluorescence level. In this sense, it is a direct measure for the initial concentration of nucleic acid in the sample. This threshold level can be set by user subjectively or it is based on an exact computing algorithm. The ***CP*** applied within this paper was obtained direct from LightCycler software version 3.5 (Roche Diagnostics) based on a special computation method (Wittwer et al.; US Patent No.: US 6,303,305 B1), where ***CP*** is placed into the positive maximum of the second derivative of the curve (Figure 1), computed on smoothed PCR fluorescence data. Thus ***CP*** denotes the second derivative maximum and x_0 is the first derivative maximum, of the described mathematical model. ***CP*** gives information similar to x_0, and is less influenced by reaction inhibitors and inhomogenities, like the PCR efficiency (parameter ***b***; Tichopad et al. 2003).
In the two-step RT PCR reaction the y_0 parameter becomes irrelevant for the analysis since it reflects only the pipetting error and the background noise of the measuring optic system of the LightCycler. In the one step approach, parameters ***a, b,*** x_0 and ***CP*** were analysed together with the y_0 parameter. Here, the higher ***y0*** the higher the cDNA synthesis efficiency of the previous RT reaction was. This parameter becomes an interesting hint to assume RT efficiency, provided, the one-step approach was employed.
All further statistics was done using the SAS 8.02 software. Effect of the flavanols was inspected for all model parameters. Comparison of samples with one of the two flavanols versus control group was carried out. The one-way ANOVA test was employed, testing, whether there are differences between the samples containing a special flavanol (each n = 12) and the control group (n = 8). The differences between dilution-steps of flavanol added into samples were disregarded in the model. The ANOVA model was reduced to the *present/absent* character. The applied measurement was taken as the dilution steps of factor 10 was too great and the assumption of dilution response would thus weaken the test's differentiation between groups with present and absent flavanol. Since the statistical design was heavily unbalanced, the procedure Generalised Linear Model (GLM) was employed. Data was plotted and visually inspected (not all figures shown). A table of p values for parameters of the sigmoidal model was established (Table 1). Statistical probability of obtaining results with $p<0.05$ was considered significant. Where significant statistics was obtained a short comment was given (Table 1).

Results

Modified extraction procedure of the total RNA from bovine samples according to Chomczynski (1993) could produce highly pure RNA extract with high integrity as inspected by electrophoresis and melting curve inspection (Ririe et al 1997). The RT as well as the PCR reaction showed a good performance and gave a typical shaped sigmoidal curve with a steep exponential phase and sudden plateau termination of the reaction, resulting in a high plateau phase (compare figure 1). Inspection of the melting curve of the amplification product showed no side products generated during the reaction (e.g. primer-dimers).

The four-parametric sigmoid model could tightly fit the fluorescence data (r^2>0.999, n = 64) and thus produced reliable model parameters for the further estimation. These parameters were subsequently compared between the control group and the positive samples, employing the one-way ANOVA test (Table 1). Before employing the one-way ANOVA test the data was checked for the Gaussian distribution and equal variance between groups. Where this condition was fulfilled computation of the one-way ANOVA test was performed directly on the data. Often, no linear trend in the model parameters corresponded to the dilution series (Figure 2) or the error distribution within the dilution groups was heavily unequal (Figure 3). For this reason no linear regression analysis or adjustment to the dilution-covariate (ANCOVA model) was performed. The figures 2 and 3 were chosen to demonstrate the difference between the control group and groups with various concentrations of the two compounds. They also show the error distribution and the pattern of the concentration effect.

The effect of EGCG on the RT reaction was statistically significant (p=0.029) and caused a lower RT product yielded (Table 1). The decrease of the y_0 value due to EGCG addition was approximately 0.2 fluorescence units. No response to the concentration gradient, in the form of a linear trend, was present on the data plot.

Other parameters reflecting an effect on the subsequent polymerase reaction remained unaffected. This is logical, as there was no longer the same cDNA concentration in samples at the beginning of the polymerase reaction. This interfered with the real effects of flavanols during the following PCR cycling. Therefore the two-step approach had to be employed to get rid of any interference with the prior RT.

In the two-step approach, both compounds caused very significant (p<0.0001) delay of reaction as reflected by later reaching of the x_0 or the ***CP*** (p<0.0001). (+)-catechin caused +0.22 shift in the x_0, that is, the reactions containing (+)-catechin reached their first derivative maximum, in average, 0.22 cycles later. Adding EGCG into reaction caused a shift of 0.17 for x_0 and 0.27 for ***CP***. (+)-Catechin caused additionally a direct efficiency decrease of the reaction as reported by the ***b*** parameter change of 0.04 (p=0.0015). This decrease of efficiency was strongly linearly concentration dependent, but the error distribution between different concentration groups was heavily unequal (Figure 3).

Discussion

We studied the inhibition of the *Thermus aquaticus (Taq)* polymerase and *Moloney Murine Leukemia Virus (MMLV)* reverse transcriptase performance in addition of two tea flavanol compounds (+)-catechin and EGCG using real-time PCR with reverse transcribed mRNA (Wittwer et al. 1997, Ririe et al. 1997, Rasmussen 2001). This experimental model with both enzymes was just a rough approximation of biological conditions in a real cell. We compared several reaction trajectories obtained with or without the addition of these two flavanol compounds.

Monitoring the reaction kinetics of real-time PCR or/and real-time RT-PCR (Wittwer et al. 1997) is a possible way to study the performance of a DNA polymerase and also reverse transcriptase under various conditions. If a suitable mathematical model is applied to fit the reaction history a quantitative analysis of the reaction kinetics is possible (Tichopad et al. 2002).

ONE-STEP approach A

	a	b	x_0	y_0	CP
CATECHIN p-value *The addition of agent causes*	0.8052 -	0.1858 -	0.9416 -	0.4086 -	0.5315 -
EGCG p-value *The addition of agent causes*	0.1051 -	0.5921 -	0.6113 -	**0.0294** *lower RT product*	0.5171 -

TWO-STEP approach B

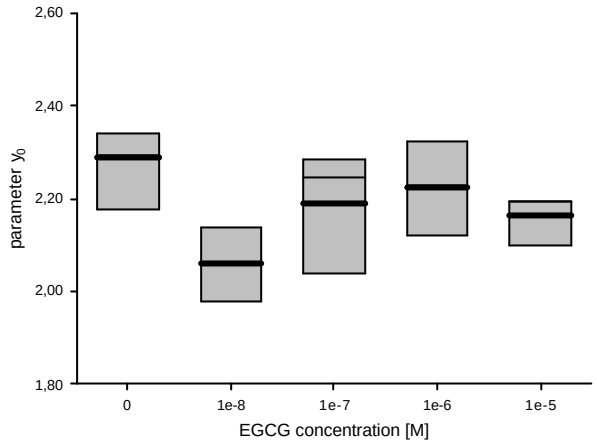
	a	b	x_0	CP
CATECHIN p-value *The addition of agent causes*	0.0815 -	**0.0015** *decrease of PCR efficiency*	**<0.0001** *delay of PCR*	0.8843 -
EGCG p-value *The addition of agent causes*	0.0919 -	0.777 -	**<0.0001** *delay of PCR*	**<0.0001** *delay of PCR*

Table 1: The significance of the alteration of parameters by EGCG and (+)-catechin in one-step and two-step real-time RT-PCR approach. ***a, b, x_0, y_0*** are the parameters of the four-parametric sigmoid model and ***CP*** is the fundamental parameter of the real-time PCR quantification. Significantly altered parameters are indicated. Beneath the respective significant p-value a comment on the impact of the finding on the reaction is given.

The one-step real-time RT-PCR approach applied here facilitated the look at effect of added agents on the performance of the RT reaction and the reverse transcriptase enzyme in particular. An effect of EGCG on the one-step RT-PCR was significantly present as decrease of final cDNA product after RT reaction. This showed that EGCG inhibits processes such as retroviral reproduction directly on the reverse transcription level. This is in accordance with reported anti-retroviral properties of EGCG and other flavanol compounds in cell and tissue cultures (Fassina et al. 2002, Tao 1992).

As the *in vitro* experiments herein were very restrictive, with relatively few reactants in comparison to complex conditions in the living cell nucleus, we can assume direct effect on the reverse transcriptase *MMLV* enzyme itself.

Employing the two-step RT-PCR approach, one can see an effect of both compounds on the polymerase reaction. With (+)-catechin, the direct efficiency and the reaction delay was altered in a sense of PCR inhibition. Also with EGCG the delay of the reaction is present and reported by change of parameters ***x_0*** and ***CP***. Different model parameters are altered by adding either of the two flavanols. This could be explained only partially. Each of the two compounds has possibly a different saturation plateau in its inhibition of the reaction. This effect may cause the kinetic curve of the PCR to have a slightly different history and therefore the model parameters to shift. In all significant results obtained, there were no alteration of the investigated parameters towards better performance of the polymerase reaction. This is a strong indication that EGCG as well as (+)-catechin inhibits or just slows down the efficiency rate of PCR. Since the reaction system is relatively simple, containing just few reaction components, we assume a direct effect on the *Taq* polymerase itself.



Figures 2. Effect of EGCG on RT reaction parameter y_0 presented in log transformation on PCR reaction parameter ***b.*** The first box represents 8 reactions with no EGCG added. Following boxes of n=3 represent reactions with various concentration of EGCG added.
The length of the box represents the interquartile range (the distance between the 25th and the 75th percentiles), the solid line interior represents the mean and the thin line in the box interior represents the median.

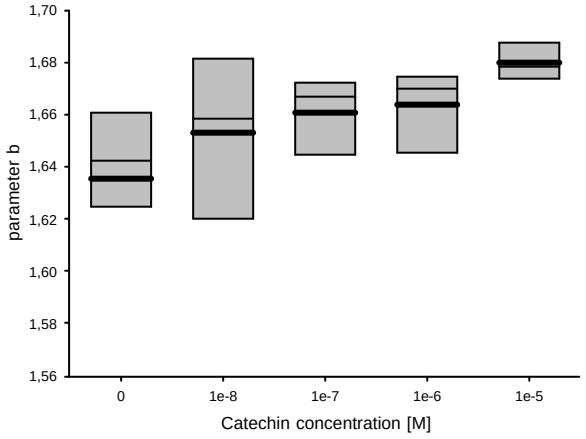


Figure 3. Effect of (+)-catechin on PCR reaction parameter ***b*** (direct reaction efficiency presented in log transformation).

The range of dilution of flavanols added to the reaction was biologically relevant as to be able to inhibit the PCR polymerase and RT reaction. Even low concentrations, down to from 10^{-5} to 10^{-8} M, exhibit significant inhibitory effect on the investigated enzymes, compared to untreated controls. Unfortunately, no quantitative conclusion can be drawn from the used flavanols concentrations.

The real-time PCR approach employed to study the performance of a polymerase reaction is not the most common but surely a very sensitive one. Unfortunately, the real-time PCR cannot be used quantitatively in this study, because the fluorescence data is directly analysed. There is always a great deal of background noise of fluorescence that is hard to quantify. Because of this background noise it is impossible to quantitatively express the difference between the treated and the control samples. However the model can detect very sensitively each increase or decrease shift in the reaction system.

The PCR can only roughly simulate the biological conditions in the eukaryotic cell. The fact, that the *Taq* polymerase is of archeobacterial and not of eukaryotic origin must be taken into account. On the other hand, the PCR model of DNA replication *in vivo* can restrict the reaction to only few compounds present, facilitating thus more robust interpretation. Surely, more tightly focused study must be performed to disclose whether flavanol compounds can decrease the risk of mutation during DNA replication by slowing the rate of enzyme reaction.

Further we can assume, that the applied PCR reaction with 45 cycles is a model for cell proliferation with 45 mitoses, where DNA is also doubled. Herein, a significant influence on polymerase activity is given. If we speculate how many cell cycles (with an average epithelium cell live span of a few days) are occurring in the gut epithelium during a human live of 75 years, we may estimate the significance of the inhibitory effect of flavanols in the gut. Gastro intestinal cell proliferation rate will be slowed down and therefore the risk of gut cancer may be markedly reduced.

To conclude, employing modern and sensitive diagnostic method, we present sensitive evidence in this paper, that the tea flavanols (+)-catechin and EGCG can inhibit and slow down the DNA

polymerase reaction and reverse transcription. This is of concern as far as anticancer properties of this compounds are discussed.

Acknowledgement
The authors thank Prof. W. Feucht from the Lehrstuhl für Obstbau of the Technical University of Munich for his initiating this paper and plenty of inspiring ideas.

References

Ahn, W.S., Huh, S.W., Bae, S.M., Lee, I.P., Lee, J.M., Namkoong, S.E., Kim, C.K. and Sin, J.I., 2003. A major constituent of green tea, EGCG, inhibits the growth of a human cervical cancer cell line, CaSki cells, through apoptosis, G(1) arrest, and regulation of gene expression. DNA Cell Biology 22, 217-24.

Cabrera, C., Gimenez, R. and Lopez, M.C., 2003. Determination of tea components with antioxidant activity. J Agric Food Chem 51, 4427-35.

Chang, L.K., Wei, T.T., Chiu, Y.F., Tung, C.P., Chuang, J.Y., Hung, S.K., Li, C. and Liu, S.T., 2003. Inhibition of Epstein-Barr virus lytic cycle by (-)-epigallocatechin gallate. Biochemical and Biophysical Research Communications 301, 1062-8.

Chomczynski, P.A., 1993. Reagent for single-step simultaneous isolation of RNA. BioTechniques 15, 532-536

Fassina, G., Buffa, A., Benelli, R., Varnier, O.E., Noonan, D.M. and Albini, A., 2002. Polyphenolic antioxidant (-)-epigallocatechin-3-gallate from green tea as a candidate anti-HIV agent. AIDS 16, 939-41.

Gupta, S., Hussain, T. and Mukhtar, H., 2003. Molecular pathway for (-)-epigallocatechin-3-gallate-induced cell cycle arrest and apoptosis of human prostate carcinoma cells. Archives of Biochemistry and Biophysics 410: 177-185.

Hayakawa, S., Saeki, K., Sazuka, Y., Shoji, Y., Ohta, T., Kaji, K., You, M. and Isemura, M., 2001. Apoptosis induction by epicatechin gallate involves its binding to FAS. Biochemical and Biophysical Research Communications 24, 1102-1106.

Henning, S.M., Fajardo-Lira, C., Lee, H.W., Youssefian, A.A., Go, V.L. and Heber D., 2003. Catechin content of 18 teas and a green tea extract supplement correlates with the antioxidant capacity. Nutr Cancer 45, 226-35.

Higdon, J.V. and Frei, B., 2003. Tea Catechins and Polyphenols: Health Effects, Metabolism and Antioxidant Functions. Critical Reviews in Food Science and Nutrition 43, 89-143.

Lambert, J.D. and Yang, C.S., 2003. Cancer chemopreventive activity and bioavailability of tea and tea polyphenols. Mutatation Research 523-524, 201-208.

Middleton, E. Jr., Kandaswami, C. and Theoharides, T.C., 2000. The effects of plant flavonoids on mammalian cells: implications for inflammation, heart disease, and cancer. Pharmacology Reviews 52, 673-751.

Morre, D.J., Morre, D.M., Sun, H., Cooper, R., Chang, J. and Janle, E.M., 2003. Tea Catechin Synergies in Inhibition of Cancer Cell Proliferation and of a Cancer Specific Cell Surface Oxidase (ECTO-NOX). Pharmacology & Toxicology 92, 234-41.

Morrison, T.B., Weis, J.J. and Wittwer, C.T., 1998. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. Biotechniques 24, 954-958.

Muto, S., Fujita, K., Yamazaki, Y., Kamataki, T., 2001. Inhibition by green tea catechins of metabolic activation of procarcinogens by

human cytochrome P450. Mutation Research 479, 197-206.

Okabe, S., Fujimoto, N., Sueoka, N., Suganuma, M. and Fujiki, H., 2001. Modulation of Gene Expression by (-)-Epigallocatechin Gallate in PC-9 Cells Using a cDNA Expression Array. Biol Pharm Bull 24, 883-886.

Polster, J., Dithmar, H. and Feucht W., 2003. Are histones the targets for flavan-3-ols (catechins) in nuclei. Biological Chemistry 384, 997-1006.

Rasmussen, R., 2001. Quantification on the LightCycler instrument. In In: S. Meuer, C. Wittwer and K. Nakagawara (Eds.), Rapid cycle real-time PCR: Methods and Applications, Springer, Heidelberg, pp. 21-34.

Ririe, K.M., Rasmussen, R.T. and Wittwer, C.T., 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Analytical Biochemistry 245, 154–60.

Seimiya, H., Oh-hara, T., Suzuki, T., Naasani, I., Shimazaki, T., Tsuchiya, K. and Tsuruo, T., 2002. Telomere shortening and growth inhibition of human cancer cells by novel synthetic telomerase inhibitors MST-312, MST-295, and MST-1991. Molecular Cancer Therapy 1, 657-65.

Schwarz, D. and Roots, I., 2003, In vitro assessment of inhibition by natural polyphenols of metabolic activation of procarcinogenes by humans CYP1A1. Biochemical and Biophysical Research Communications 303, 902-907.

Tao, P., 1992, [The inhibitory effects of catechin derivatives on the activities of human immunodeficiency virus reverse transcriptase and DNA polymerases.] Zhongguo Yi Xue Ke Xue Yuan Xue Bao. 14, 334-8.

Tichopad, A., Dilger, M., Schwarz, G. and Pfaffl, M.W., 2003. Standardized determination of real-time PCR efficiency from a single reaction set-up. Nucleic Acids Research. 31, E122.

Tichopad, A., Dzidic, A., Pfaffl, M.W., 2002. Improving quantitative real–time RT–PCR reproducibility by boosting primer–linked amplification efficiency. Biotechnology Letters 24, 2053-2056.

Wiseman, S.A., Balentine, D.A. and Frei, B., 1997. Antioxidants in tea. Critical Reviews in Food Science and Nutrition 37, 705-718.

Wittmann, S.L., Pfaffl, M.W. and Meyer, H.H.D. and Bruckmaier, R.M., 2002. 5-Lipoxygenase, Cyclo-oxygenase-2 and Tumor necrosis factor alpha (TNF-alpha) gene expression in somatic milk cells. Milk Science International 57, 63-66.

Wittwer, C.T., Gutekunst, M. and Lohmann, S. Method for quantification of an analyte. United States Patent No.: US 6,303,305 B1.

Wittwer, C.T., Ririe, K.M., Andrew, R.V., David, D.A., Gundry, R.A. and Balis, U.J., 1997. The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. Biotechniques 22, 176–81.

Yamaguchi, K., Honda, M., Ikigai, H., Hara, Y. and Shimamura, T., 2002. Inhibitory effects of (-)-epigallocatechin gallate on the life cycle of human immunodeficiency virus type 1 (HIV-1). Antiviral Research 53, 19-34.

Distribution-insensitive rank-order dissimilarity measure based clustering on real-time PCR gene expression data of steadily expressed standards

Aleš Tichopád[1,2], Ladislav Pecen[1], Michael W. Pfaffl[2]

[1]*IMFORM GmBH, International Clinical Research, Darmstadt, Germany*

[2]*Lehrstuhl für Physiologie, Fakultät Wissenschaftszentrum Weihenstephan, Technische Universität München, Freising-Weihenstephan, Germany*

***Corresponding author**: Aleš Tichopád, IMFORM GmBH, Birkenweg 14, Darmstadt, Germany*

ABSTRACT

Cluster analysis is a tool often employed in the micro-array techniques but less in the real-time PCR. Herein, instead of the Euclidian distances correlation coefficient is taken as a dissimilarity measure. The dissimilarity measure is made robust using a rank-order correlation coefficient rather then a parametric one. There is no need for an overall probability adjustment as in scoring methods based on repeated pair-wise comparisons. The rank-order correlation matrix gives a good base for clustering procedure of gene expression data obtained by real-time RT-PCR as it disregards the different expression levels. Associated with each cluster is a linear combination of the variables in the cluster, which is the first principal component. Large set of variables can often be replaced by the set of cluster components with little loss of information. In this way, distinct clusters containing unregulated housekeeping genes along with other steadily behaved genes can be disclosed and utilized for standardization purposes. Simple SAS macro was written to facilitate the computing procedure. Dummy data together with data from biological experiment were taken to validate the method. In both cases good intuitive results were obtained.

BACKGROUND

Search for genes unregulated under treatment is an essential task before any relative gene-expression quantification is conducted. Where a change due to treatment in studied gene's expression is measured, reference gene/s must be employed. The reference transcriptome is assumed to remain constant in abundance under applied treatment, and any change in it can be assigned to assay disturbances. The same disturbances are then expected to affect the results of the studied gene. This is the principle of relative gene quantification with reference gene employed. Computing method was proposed that incorporates the reference data (Pfaffl, 2001). Unlike the highly elaborated mathematical aspects, the biological aspects of the problematic remain unclear. Papers reporting about regulation of assumed housekeeping genes are published to often to believe that there are some perfectly unregulated housekeeping genes (Thellin et al., 1999; Schmittgen et al., 2000). Physiological changes in untreated organism can, alone, cause regulation of these genes (Yamada et al., 1997). Vandesompele et al. (2002) proposed a computing method based on the standard deviation that orders the candidates according to their best pair-wise score with other genes. This method, however, does not reflect the target genes and their relation to the reference genes. Repeated pair-ways analysis on more then two genes (Pfaffl et al. 2004) is confronted with the need to adjust the overall probability value.

Some simple approach ignoring the imaginary boundary between unregulated housekeeping genes and regulated genes is desired, that would group genes, based on a robust distribution-insensitive dissimilarity measure. Spearman rank-order correlation coefficient is a nonparametric measure of association based on the rank of the data values. The formula is

$$q = \frac{\Sigma(R_i - R)(S_i - S)}{\sqrt{\Sigma(R_i - \overline{R})^2 \Sigma(S_i - \overline{S})^2}}$$

where R_i is the rank of the *i-th* x value, S_i is the rank of the *i-th y* value, $\overline{R}$ is the mean of the R_i values, and $\overline{S}$ is the mean of the S_i values.

Clustering procedure based on the Spearman correlation coefficient prevents the erroneous results due to non-normal distributed real-time PCR data (Urban et al. 2003). In here proposed method, associated with each cluster is a linear combination of the variables in the cluster, which is the first principal component. A large set of variables can often be replaced by the set of cluster components with little loss of information. A given number of cluster components does not generally explain as much variance as the same number of principal components on the full set of variables, but the cluster components are usually easier to interpret than the principal components. The first principal component is a weighted average of the variables that explains as much variance as possible. Principal components have a variety of useful properties (Rao 1964; Kshirsagar 1972):

The eigenvectors are orthogonal, so the principal components represent jointly perpendicular directions through the space of the original variables.

The principal component scores are jointly uncorrelated. Note that this property is quite distinct from the previous one.

The first principal component has the largest variance of any unit-length linear combination of the observed variables. The jth principal component has the largest variance of any unit-length linear combination orthogonal to the first j-1 principal components. The last principal component has the smallest variance of

any linear combination of the original variables.

The scores on the first j principal components have the highest possible generalized variance of any set of unit-length linear combinations of the original variables.

The first j principal components provide a least-squares solution to the model

Y = XB + E

where **Y** is an $n \times p$ matrix of the centered observed variables; **X** is the $n \times j$ matrix of scores on the first j principal components; **B** is the $j \times p$ matrix of eigenvectors; **E** is an $n \times p$ matrix of residuals; and the trace(**E**'**E**), the sum of all the squared elements in **E**, is to be minimized. In other words, the first j principal components are the best linear predictors of the original variables among all possible sets of j variables, although any nonsingular linear transformation of the first j principal components would provide equally good prediction. The same result is obtained if the determinant or the Euclidean norm of **E**'**E** rather than the trace is to be minimized.

In geometric terms, the j-dimensional linear subspace spanned by the first j principal components provides the best possible fit to the data points as measured by the sum of squared perpendicular distances from each data point to the subspace. This is in contrast to the geometric interpretation of least squares regression, which minimizes the sum of squared vertical distances. For example, suppose you have two variables. Then, the first principal component minimizes the sum of squared perpendicular distances from the points to the first principal axis. This is in contrast to least squares, which would minimize the sum of squared vertical distances from the points to the fitted line.

MATERIALS AND METHODS

Dummy data set

For this data set, one biological variable was taken and three dummy variables were created, each of n=30. First variable called RG contains crossing point data from real gene quantification assay (Ubiquitin).

$$RG \sim U(\mu_{RG}, s^2_{RG})$$

Second and third semi-random dummy variable SRD1 and SRD2 were created as follows:

$$SRD_1 = 1.3*RG + R_1, \quad R_1 \sim Z(\mu=0, s^2=1)$$

$$SRD_2 = 1.1*RG + R_2, \quad R_2 \sim N(\mu=0, s^2=4)$$

where RG is the real gene's crossing point, R_1 is a number randomly generated from the standardized normal distribution Z (i.e. with mean value $\mu = 0$ and standard deviation $s^2 = 1$) and R_2 is a random number generated from the normal distribution with mean value $\mu = 0$ and standard deviation $s^2 = 4$. These two random additive exponents R_1 and R_2 introduce some disturbance into, otherwise perfect, correlation between SRD_1 or SRD_2 and RG. The SRDs differ one from another not only by the linear exponent but also by the standard deviation of the random additive increment (R_1, R_2). Therefore, the SRD_2 is supposed to be less correlated with the RG then the SRD_1. The SRD_1 should be better correlated with RG but its values should be more distanced (in Euclidian sense) since its linear coefficient is greater then in SRD_2.

The fourth random dummy variable called RD is generated as a random number from the normal distribution with the same mean and standard deviation as the RG.

$$RD \sim N(\mu_{RG}, s^2_{RG})$$

After debugging the SAS code generating the three variables, the data of the first functional run was taken to prevent subjective decision on data suitability.

Descriptive statistics on the data obtained is shown in table 1.

Gene	Mean	Variance	Std Dev
RG	20.85	1.05	1.02
SRD1	27.31	3.45	1.85
SRD2	22.74	2.72	1.65
RD	21.00	1.04	1.02

Table 1. Descriptive statistics on the dummy data.

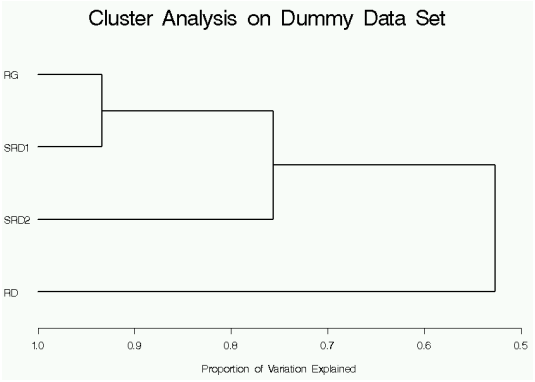


Figure 1. Cluster analysis on the dummy data.
RG denotes the real gene data (n=30). $RG \sim U(\mu_{RG}, s^2_{RG})$
The following two semi-random dummy variables are computed out of the RG as follows:
$SRD1 = 1.3*RG + R1 \quad R1 \sim Z(\mu = 0, s2 = 1)$

$SRD2 = 1.1* RG + R2 \quad R2 \sim N(\mu = 0, s2 = 4)$
RD denotes random dummy random variable generated from normal distribution with the same parameters like the RG. $RD \sim N(\mu_{RG}, s^2_{RG})$

Biological data set

Total RNA from 31 bovine Corpora Lutea samples was extracted from small slices of deep frozen CL with peqGOLD according to the manufacturer's instruction (Chomczynski, 1993).
The cDNA was reverse-transcribed from 1000 ng total RNA with 2000 units of M-MLV Reverse Transcriptase (Promega corp., Madison, USA) according to the manufacturer instructions.
Data on expression levels of studied factors were obtained on LightCycler (Roche, Basel, Switzerland) PCR instrument (Wittwer et al., 1997; Rasmussen, 2001). In the 31 cDNA samples expression of four genes with assumed stable expression – housekeeping genes (HKG); Ubiquitin (UBQ), Glyceraldehyd-3-Phosphate Dehydrogenase (GAPD), ß-actin and 18S ribosomal unit was quantified together with ten studied target genes; IGF-1 (insulin-like growth factors type 1), IGF-2, IGFR-1 (Insulin-like growth factor receptor type 1), IGFR-2, IGFBP-1 (Insulin-like growth factor binding protein type 1) – IGF-6, those expression is studied. In each biological sample all 14 factors were quantified. Descriptive statistics computed on this data is shown in table 2.
The data was analysed by following SAS macro:

```
data genes;
   input UBQ GAPD   Beta-actin      S18
   IGF-1...;
   cards;
20.59  21.06  17.80  10.00  28.49  .  .  .
21.17  20.84  18.14  12.92  29.05  .  .  .
20.67  20.09  17.84  9.87   29.69  .  .  .
20.99  20.78  18.30  10.15  28.74  .  .  .
19.77  19.65  16.71  11.66  29.03  .  .  .
19.91  21.33  17.22  10.37  27.59  .  .  .
20.75  21.74  17.58  10.05  28.97  .  .  .
21.08  21.25  17.16  13.03  28.51  .  .  .
19.22  21.24  17.44  12.58  28.87  .  .  .
.  .  .  .  .  .  .  .
.  .  .  .  .  .  .  .
.  .  .  .  .  .  .  .
run;

%macro CLUSTER (var, clus);

   PROC CORR outs=cor;
      var &VAR;

   PROC VARCLUS data=cor outtree=tree
maxclusters=&clus;
      var &VAR;

   axis2 minor=none;
   axis1     label=('Proportion     of
Variation Explained') minor=none;

   PROC  TREE  horizontal  vaxis=axis2
haxis=axis1 lines=(width=2);
      height _propor_;
   run;
```

```
%mend cluster;

%cluster (var=UBQ GAPD Betaactin S18
IGF1...,  clus=14);
```

The CORR procedure is a statistical procedure that computes Spearman correlation coefficient. The correlation matrix is then saved as an output data set *cor*.
The PROC VARCLUS statement starts the VARCLUS procedure. The procedure uses the recently created *cor* data set and omits observations with missing values from the analysis. The MAXCLUSTERS= option specifies the largest number of clusters desired. This can be determined by macro invocation as *&clusno* parameter. The VARCLUS procedure tries to maximize the sum across clusters of the variance of the original variables that is explained by the cluster components. Either the correlation or the covariance matrix can be analyzed. The set of variables is divided into nonoverlapping clusters in such a way that each cluster can be interpreted as essentially unidimensional. For each cluster, PROC VARCLUS computes a component that is the first principal component and tries to maximize the sum across clusters of the variation accounted for by the cluster components. PROC VARCLUS is a type of oblique component analysis related to multiple group factor analysis (Harman 1976). By default, PROC VARCLUS begins with all variables in a single cluster. It then repeats the following steps:

1. A cluster is chosen for splitting.

2. The chosen cluster is split into two clusters by finding the first two principal components, performing an orthoblique rotation (raw quartimax rotation on the eigenvectors), and assigning each variable to the rotated component with which it has the higher squared correlation.

3. Variables are iteratively reassigned to clusters to maximize the variance accounted for by the cluster components. The reassignment may be required to maintain a hierarchical structure.

By default, PROC VARCLUS stops when each cluster has only a single eigenvalue greater than one, thus satisfying the most popular criterion for determining the sufficiency of a single underlying factor dimension. The iterative reassignment of variables to clusters proceeds in two phases. The first is a nearest component sorting (NCS) phase, similar in principle to the nearest centroid sorting algorithms described by Anderberg (1973). In each iteration, the cluster components are computed, and each variable is assigned to the component with which it has the highest squared correlation. The second phase involves a search algorithm in which each variable is tested to see if assigning it to a different cluster increases the amount of variance explained. If a variable is reassigned during the search phase, the components of the two clusters involved are recomputed before the next variable is tested. The NCS phase is much faster than the search phase but is more likely to be trapped by a local optimum. The OUTTREE= option creates an output data set to contain information on the tree structure that can be used by the TREE procedure to print a tree diagram. PROC VARCLUS displays a cluster summary and a cluster listing (table 3). The cluster summary gives the number of variables in each cluster and the variation explained by the cluster component. The proportion of variance explained is obtained by dividing the variance explained by the total variance of variables in the cluster. If the cluster contains two or more variables the second largest eigenvalue of the cluster is also printed. The cluster listing gives the variables in each cluster. Two squared correlations are calculated for each cluster. The column labeled "Own Cluster" gives the squared correlation of the variable with its own cluster component. This value should be higher than the squared

correlation with any other cluster unless an iteration limit has been exceeded. The larger the squared correlation is, the better. The column labeled "Next Closest" contains the next highest squared correlation of the variable with a cluster component. This value is low if the clusters are well separated. The column headed "1 -R^2 Ratio" gives the ratio of one minus the "Own Cluster" R^2 to one minus the "Next Closest" R^2. A small "1 –R^2 Ratio" indicates a good clustering.

The TREE procedure produces a horizontally oriented tree diagram, also known as a dendrogram or phenogram, using a data set created by the VARCLUS procedure. The AXIS statements create AXIS definitions that specify the characteristics of an axis. From left to right in the diagram, objects and clusters are progressively joined until a single, all-encompassing cluster is formed at the right (or root) of the diagram. Clusters exist at each level of the diagram, and every vertical line connects leaves and branches into progressively larger clusters (figure 1, 2 and 3). The macro is terminated by the %mend cluster. Invocation of the macro consists of the %cluster sentence and the definitions of the three macro parameters for the names of genes analyzed (*var*) and number of clusters (*clus*).

RESULTS

Results from the dummy data

The dummy data taken herein for analysis provided a good intuitive model of more and less correlated gene-expression data. The cluster diagram produced corresponded with the expected result, although the random shift was introduced into the data. The closest association was found between the RG and the SRD1 as the linear member 1.3 is greater then in SRD2 and the random variable R1 has a variance of ‘only’ 1. Following SRD2 also showed an association with the RG-SRD1 as there still was a linear relation given by the 1.1. The entirely random RD was not associated with any variable at all.
In repeated program runs, slightly differing results were obtained as the random values were newly generated, nevertheless, the hierarchy of the diagram remained unchanged.

Results from the biological data

Expression data based on the Crossing Point value were obtained from LightCycler software (Roche) and descriptive statistic was computed on this data (table 1).
Taking a look at the diagram, some 6 discrete clusters, as named in the first column of the table 2, come to the fore (figure 2). The next macro-run is therefore launched creating 6 clusters (figure 3).
Clusters 1 and 6 show a great explaining power and both contain always two housekeeping genes each. There are three ways how to deal with this result.
First, the cluster 1, considered the best separated, will be taken with its all four components for standardisation purposes. Also the cluster 6 can still be considered well separated and useful for the standardisation purposes. Both the cluster 1 and the cluster 2 contain some known ‘conservative’ housekeeping genes. Alternatively, the clusters 1 and 6 will be joined, running the analysis once more, but with cluster number set 5 (results not shown). The encompassing cluster contains all the known housekeeping genes UBQ, Beta-actin, GAPD and 18S together with IGF-2, IGF-1R, BP-3 and BP-4. These four genes can be assumed unregulated and can be taken as quantification standards.
Further, some deeper insight into the regulation paterns of the target genes can be acquired from the figures 2 and 3. Surely there is no association between above proposed standards and IGF-2R, BP-5, BP-1, BP-2, BP-6 and IGF-1. These target genes can be well standardised by the genes from clusters 1 and 6.

Variable	Mean	Variance	Std Dev	Coeff of Variation
UBQ	20.8561290	1.0509578	1.0251624	4.9154009
GAPD	21.5045161	1.0456856	1.0225877	4.7552230
Betaactin	18.2861290	1.0423378	1.0209495	5.5831909
S18	12.9716129	3.6588673	1.9128166	14.7461739
IGF1	29.3093548	1.0038062	1.0019013	3.4183670
IGF2	23.1419355	1.1726028	1.0828679	4.6792450
IGF1R	24.5858065	1.1245052	1.0604269	4.3131669
IGF2R	37.8932258	0.6983826	0.8356929	2.2053886
BP1	29.3790323	9.1035224	3.0172044	10.2699244
BP2	30.5303226	1.0411632	1.0203741	3.3421660
BP3	30.0009677	3.3899157	1.8411724	6.1370433
BP4	31.1264516	2.1838170	1.4777743	4.7476479
BP5	26.7409677	2.2443424	1.4981129	5.6023138
BP6	30.3590323	2.1222157	1.4567827	4.7985148

Table 1. Descriptive statistics on the biological data.
First the cluster analysis with 14 clusters was run to disclose the entire structure (figure 1).

		R-squared with		
Cluster	Variable	Own Cluster	Next Closest	$1-R^2$ Ratio
Cluster 1	UBQ	0.6362	0.3323	0.5448
	Betaactin	0.7239	0.3441	0.4210
	IGF2	0.7547	0.5038	0.4944
	BP4	0.8092	0.7010	0.6381
Cluster 2	BP2	0.7111	0.0226	0.2956
	BP6	0.7111	0.1439	0.3374
Cluster 3	IGF2R	0.6429	0.0174	0.3635
	BP5	0.6429	0.1200	0.4059
Cluster 4	IGF1	1.0000	0.0921	0.0000
Cluster 5	BP1	1.0000	0.0225	0.0000
Cluster 6	GAPD	0.6846	0.6351	0.8645
	S18	0.6217	0.4628	0.7043
	IGF1R	0.7732	0.3552	0.3517
	BP3	0.7960	0.4084	0.3449

Table 2. Cluster listing for 6 clusters computed on the biological data. It shows how the variables are clustered. The first cluster represents UBQ, Beta-actin, IGF-2 and BP- 4, the second duster contains the BP-2 and BP-6 and so on. It also displays the R^2 value of each variable with its own cluster and the R^2 value with its nearest cluster. The R2 value for a variable with the nearest cluster should be low if the clusters are well separated. The last column displays the ratio of $1\text{-}R_{own}^2/1\text{-}R_{nearest}^2$ for each variable. Small values of this ratio indicate good clustering.

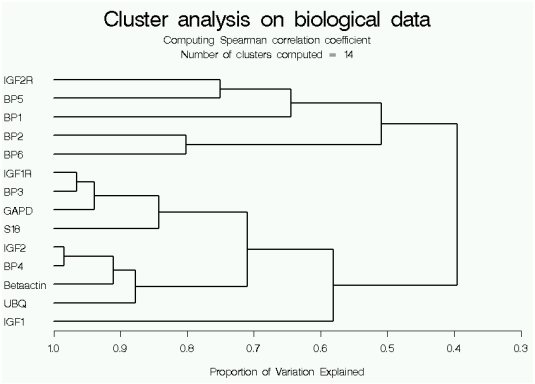


Figure 2. Cluster analysis with 14 clusters computed on the biological data.

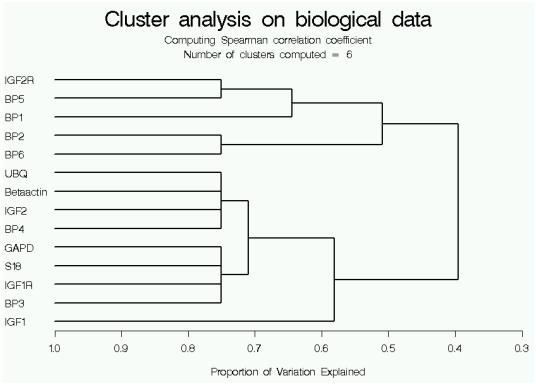


Figure 3. Cluster analysis with 6 clusters computed on the biological data

DISCUSSION

Clustering approaches have been frequently adopted on micro-array data to disclose families of co-regulated genes (Bickel, 2003; Cherepinsky et al., 2003; Raychaudhuri et al., 2001). A similar pattern of expression indicates co-regulated genes. Some genes, however, can remain untouched by the experiment.

If more such unregulated genes are compared, they as well, show a similar pattern.

This similarity is given by the stabile expression ratio between any two of the genes. Therefore, a high correlation coefficient between two unregulated genes indicates similarity. Where the sampling, extraction procedure, RT reaction, storage and the PCR performance was affected by error, all genes achieve some common erroneous shift. This shift produces some common visible pattern only in genes that are not biologically regulated, because any biological regulation would mask the minor erroneous pattern.

The success of the method of cluster analysis depends on how well its underlying model describes the patterns of expression. Based on above idea, the herein suggested cluster analysis associates genes based on similar rank-order correlation patterns as given by the correlation matrix. Genes with different expression levels but correlating well due to steady expression ratio are clustered together. The Euclidean distances cannot be taken as a measure of dissimilarities here, because the levels of expression can be different.

The real-time PCR yields so called crossing points or threshold cycles, those are the fundamental quantitative units (Rasmussen, 2001). This data shows skewed distribution and heterogeneous variance. The Gaussian distribution is only rarely given (Urban et al., 2003), therefore, the here proposed method clusters genes based on the non-parametric Spearman correlation coefficient, making the method distribution-insensitive (Bickel, 2003).

Method of stabile gene selection suggested by Vandesompele et al. (2002) can only analyse assumed independent genes. Possible co-regulation of some of the candidate housekeeping genes would bear a confounding effect into the analysis.

The clustering of the biological data was limited to 6 clusters here (figure 3). The decision on the cluster size is a trade-off between the strength of the associations within the cluster and the number of reference genes wanted. However, to see the entire association structure, the number of clusters equal to number of genes analysed is suggested (figure 2). Alternatively, algorithm deriving the cluster number from the biological background was also proposed (Bickel, 2003).

The look at the associations between all genes as facilitated by the cluster analysis can preventively exclude standardising with associated gene. If a distinct cluster contains predominantly known housekeeping genes, its genes can be applied for standardisation purposes in form of geometric mean as follows.

$$\text{Index} = \sqrt[n]{CP_1 \times CP_2 \times CP_3 \times \ldots\ldots \times CP_n}$$

where 1,2...n are the genes (Pfaffl et al. 2004). Also genes, not *a priori* assumed to be unregulated, but those were tightly clustered with housekeeping genes can be included in the index. Standardisation model for relative quantification of expression change was described by Pfaffl (2001) and Excel based spreadsheet is available Pfaffl et al. (2002).

The presented SAS macro performs the simplest mostly default computing procedures. With some knowledge of SAS, it can be modified and adapted to perform with another procedures or to produce more detailed output.

REFERENCES

Anderberg, M.R. (1973), Cluster Analysis for Applications, New York: Academic Press, Inc.

Bickel, .DR. (2003). Robust cluster analysis of microarray gene expression data with the number of clusters determined biologically. Bioinformatics. 19, 818-824.

Cherepinsky, V., Feng, J., Rejali, M., Mishra, B. (2003) Shrinkage-based similarity metric for cluster analysis of microarray data. Proc Natl Acad Sci USA. 100, 9668-9673.

Chomczynski, P. A. (1993). Reagent for the single−step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. Biotechniques 15, 532–4.

Harman, H.H. (1976), Modern Factor Analysis, Third Edition, Chicago: University of Chicago Press.

Kshirsagar, A.M. (1972), Multivariate Analysis, New York: Marcel Dekker, Inc.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in Real–time RT–PCR. Nucleic Acids Res. 1, e45.

Pfaffl, M. W., Horgan, G. W. & Dempfle, L. (2002). Relative Expression Software Tool (REST©) for group wise comparison and statistical analysis of relative expression results in Real–time PCR. Nucleic Acids Res. 30, e36.

Pfaffl, M.W., Tichopád, A., Prgomet, Ch. & Neuvians, T. (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel spreadsheet tool using a Repeated Pair-wise Correlation and Regression Analysis, Biotechnology Letters, (in Press).

Rao, C.R. The Use and Interpretation of Principal Component Analysis in Applied Research. Sankhya A 26 , 329 - 358 (1964).

Rasmussen, R. (2001) Quantification on the LightCycler instrument. In: Meuer, S., Wittwer, C., and Nakagawara, K. (eds.), Rapid cycle real - time PCR: Methods and Applications. Springer Press, Heidelberg, pp. 21-34.

Raychaudhuri, S., Sutphin, P.D., Chang, J.T. & Altman, R.B. (2001) Basic microarray analysis: grouping and feature reduction. Trends Biotechnol 19, 189-193.

Schmittgen, T. D. and Zakrajsek, B. A. (2000). Effect of experimental treatment on housekeeping gene expression: validation by real–time, quantitative RT–PCR. J. Biochem. Biophys. Methods 46, 69–81.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A. & Heinen, E. (1999). Housekeeping genes as internal standards: use and limits. J. Biotechnol. 75, 291-295.

Urban, C., Schweinberger, A., Kundi, M., Dorner, F., Hammerle, T. (2003). Relationship between detection limit and bias of accuracy of quantification of RNA by RT-PCR. Mol Cell Probes. 17, 171-4.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. & Speleman, F. (2002). Accurate normalisation of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Gen. Biol. 3, 1-12.

Wittwer, C. T., Ririe, K. M., Andrew, R. V., David, D. A., Gundry, R. A. & Balis, U. J. (1997). The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. Biotechniques 22, 176–81.

Yamada, H., Chen, D., Monstein, H-J., & Håkansen, R. (1997). Effect of Fasting on the Expression of Gastrin, Cholecystokinin, and Somatostatin Genes and of Various Housekeeping Genes in the Pancreas and Upper Digestive Tract of Rats. Biochem. Biophys. Res. Com. 231, 835-838.

Posters

Inhibition of *Taq* Polymerase and *MMLV* Reverse Transcriptase performance in presence of polyphenolic compounds: (+)-Catechine and Epigallocatechin Gallate (EGCG)

Tichopad Ales[1], Polster Jürgen[2] & Pfaffl Michael W.[1]

[1]Physiology - Weihenstephan, [2]Institute of Biological Chemistry, Technische Universität München, 85354 Freising-Weihenstephan, Germany, pfaffl@wzw.tum.de

BACKGRUND

Polyphenolic compounds present in many foods are known to have a preventive but also curative effect on carcinogenic progression by multiple effects on the cell physiology. A direct effect of polyphenols on the enzyme activity should therefore be considered.

GOAL

Herein we studied *in vitro* the inhibitory effects of two polyphenolic compounds (+)-Catechine & Epigallocatechin Gallate (EGCG) on the performance of the polymerase and reverse transcriptase, as a model for eukaryotic and viral enzyme activity.

MATERIAL & METHODS

Since in real-time RT-PCR the reaction kinetics trajectory can be recorded, we compared several amplification histories obtained with or without polyphenols.

Two different approaches of RT-PCR were adopted:

1. **A one-step RT-PCR approach** (RT and PCR together in one run), where y_0 is showing the efficiency of the prior RT reaction;
2. **A two-step RT-PCR approach**, where the mRNA was separately reverse transcribed, and polyphenols were added only into PCR;

In each approach, reaction setups without any additional agent as a background control (n = 8), and three serial dilutions of both polyphenols were performed: $1*10^{-5}$, $1*10^{-6}$, $1*10^{-7}$ and $1*10^{-8}$ M (n = 3). We determined various parameters describing the enzyme properties derived from the sigmoidal shaped reaction trajectory, using an established four parametric sigmoid model (Tichopad et al., *Biotech. Lett.* 2002; *Mol. & Cel. Probes* – 2003, in press).

$$f(x) = y_0 + \frac{a}{1 + e^{-(\frac{x - x_0}{b})}}$$

four parametric sigmoidal model

Raw fluorescence data were fitted, where ***f(x)*** is the function computed fluorescence at cycle ***x***, y_0 is the background fluorescence, ***a*** the plateau height ($a = y_{max} - y_0$), ***e*** is the natural logarithm base, ***x*** is the cycle number, x_0 is the first derivative maximum (**FDM**), the second derivative maximum (**SDM**), and ***b*** describes the slope at x_0, representing an „inverse estimator" of the polymerase efficiency. Further the area under the melting curve peak (**AUC**) of a final PCR product was determined, representing the amount of amplified product. All statistics were done in SAS 8.02 using GLM, checking for differences between the groups.

RESULTS

In one-step RT-PCR, only the effect of EGCG addition was significantly present as a decrease of final cDNA product after RT reaction (Table 1). This is in accordance with known antiviral properties of EGCG. Decrease in PCR product was a consequence of decreased prior template cDNA.

Employing two-step RT-PCR approach one can see in table 1 an effect of both compounds on PCR performance. Parameters were altered in a sense of PCR inhibition and lower PCR efficiency. The range of added polyphenols was biologically relevant (10 nM to 10 µM) and able to inhibit the enzyme activities.

CONCLUSION

Our results suggest that polyphenols are suppressing the polymerase as well as reverse transcriptase activity *in vitro*. This may lead to the hypothesis, that organs exposed to polyphenols exhibit lower DNA replication and proliferation rate, as well as lower viral activity caused by retroviruses.

Further information: http://www.wzw.tum.de/gene-quantification/

Table 1: The significance of the alteration of parameters by EGCG and (+)-Catechin.

a, b, x0 (= FDM), y0, CP and **SDM** are the parameters of the four-parametric sigmoid model. Significantly altered parameters are indicated. The lower field contains comment on the impact of the finding on the reaction. ***AUC*** represents the amount of amplified real-time PCR product.

One-step real-time RT-PCR approach

	a	b	x0 = FDM	y0	SDM	AUC
p value	0.8052	0.1858	0.9416	0.4086	0.5315	0.9273
addition of **Catechin** caused	-	-	-	-	-	-
p value	0.1051	0.5921	0.6113	***0.0294***	0.5171	***0.0462***
addition of **EGCG** caused	-	-	-	*lower RT product*	-	*lower PCR product*

Two-step real-time RT-PCR approach

p value	0.0815	***0.0015***	***<0.0001***	-	0.8843	0.4303
addition of **Catechin** caused	-	*decrease of efficiency*	*delay of PCR*	-	-	-
p value	0.0919	0.777	***<0.0001***	-	***<0.0001***	0.4090
addition of **EGCG** caused	-	-	*delay of PCR*	-	*delay of PCR*	-

Figure 1: Effect of EGCG on RT reaction efficiency (parameter ***y0***).

Figure 2: Effect of (+)-Catechin on real-time PCR reaction efficiency (parameter ***b***).

Box Plot: The first box represents 8 reactions with no EGCG added. Following boxes represent 3 reactions with various concentration of polyphenols added. The length of the box represents the inter-quartile range (the distance between the 25th and the 75th percentiles), the dot in the box interior represents the mean, the horizontal line in the box interior represents the median, the vertical lines issuing from the box extend to the minimum and maximum values of the analysis variable.

Fig. 1

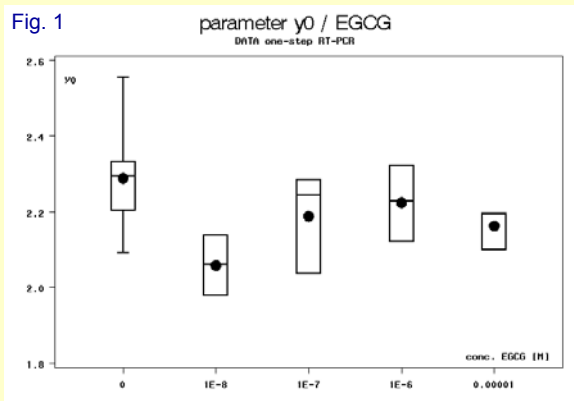
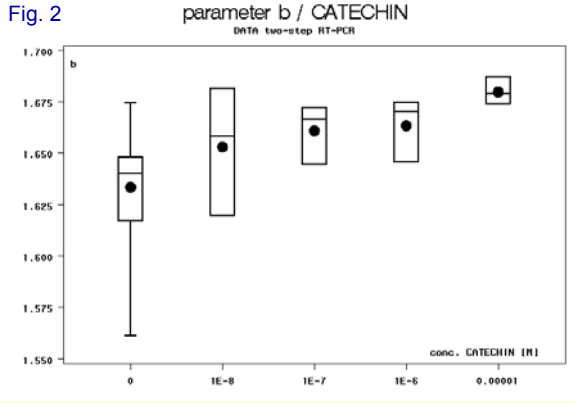


Fig. 2



Search by cluster analysis for steadily expressed genes with application as normalization index in real-time RT-PCR

Ales Tichopad[1]& Michael W. Pfaffl[2]

[1]IMFORM GmbH. International Clinical Research, Birkenweg 14, D-94295 Darmstadt; [2]Physiology - Weihenstephan, Zentralinstitut für Ernährung- und Lebensmittel-forschung, Technische Universität München, 85354, Freising-Weihenstephan

BACKGROUND

Search for genes unregulated under treatment is an essential task before any relative gene-expression quantification can be conducted. Some simple approach ignoring the imaginary boundary between unregulated housekeeping genes and regulated genes is desired, that would group genes, based on a robust distribution-insensitive similarity measure.

MATERIAL & METHODS

Total RNA from 31 bovine Corpora Lutea was extracted. Data on expression levels of studied factors were obtained on LightCycler. In the 31 cDNA samples expression of four genes with assumed stable expression – housekeeping genes (HKG); Ubiquitin (UBQ), Glyceraldehyd-3-Phosphate Dehydrogenase (GAPD), ß-actin and 18S ribosomal unit was quantified together with ten studied target genes; IGF-1 (insulin-like growth factors type 1), IGF-2, IGFR-1 (Insulin-like growth factor receptor type 1), IGFR-2, IGFBP-1 (Insulin-like growth factor binding protein type 1) – IGF-6, those expression is studied.

Similarity measure computation

Spearman rank-order correlation coefficient is a nonparametric measure of association based on the rank of the data values. The formula is

$$q = \frac{\Sigma(R_i - R)(S_i - S)}{\sqrt{\Sigma(R_i - \overline{R})^2 \Sigma(S_i - \overline{S})^2}}$$

where R_i is the rank of the *i-th* x value, S_i is the rank of the *i-th y* value, $\overline{R}$ is the mean of the R_i values, and $\overline{S}$ is the mean of the S_i values.

Clustering procedure based on the Spearman correlation coefficient prevents the erroneous results due to non-normal distributed real-time PCR data.

Clustering procedure

Associated with each cluster is a linear combination of the genes in the cluster, which is the first principal component. A large set of genes can often be replaced by the set of cluster components with little loss of information. The first j principal components provide a least-squares solution to the model

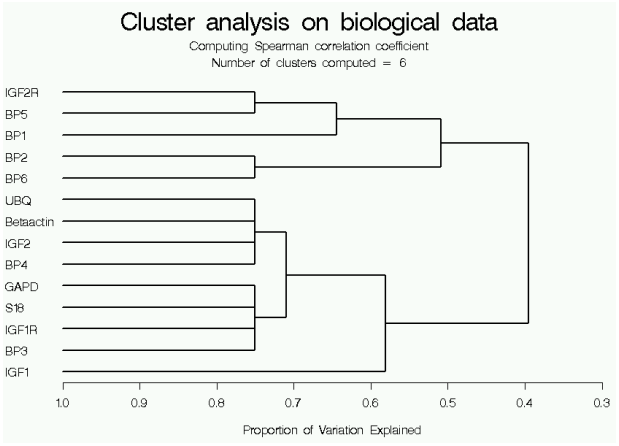
$$\mathbf{Y = XB + E}$$

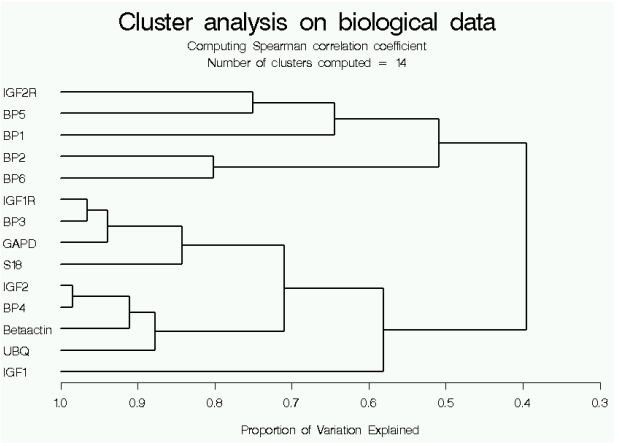
where **Y** is an $n \times p$ matrix of the centered observed variables; **X** is the $n \times j$ matrix of scores on the first j principal components; **B** is the $j \times p$ matrix of eigenvectors; **E** is an $n \times p$ matrix of residuals; and the trace(**E'E**), the sum of all the squared elements in **E**, is to be minimized.

			R-squared with	
		Own	Next	1-R²
Cluster	Variable	Cluster	Closest	Ratio
Cluster 1	UBQ	0.6362	0.3323	0.5448
	Betaactin	0.7239	0.3441	0.4210
	IGF2	0.7547	0.5038	0.4944
	BP4	0.8092	0.7010	0.6381
Cluster 2	BP2	0.7111	0.0226	0.2956
	BP6	0.7111	0.1439	0.3374
Cluster 3	IGF2R	0.6429	0.0174	0.3635
	BP5	0.6429	0.1200	0.4059
Cluster 4	IGF1	1.0000	0.0921	0.0000
Cluster 5	BP1	1.0000	0.0225	0.0000
Cluster 6	GAPD	0.6846	0.6351	0.8645
	S18	0.6217	0.4628	0.7043
	IGF1R	0.7732	0.3552	0.3517
	BP3	0.7960	0.4084	0.3449

TAB 1. Cluster listing. It displays the R^2 value of each variable with its own cluster and the R^2 value with its nearest cluster. The R2 value for a variable with the nearest cluster should be low if the clusters are well separated. The last column displays the ratio of $1\text{-}R_{own}^2/1\text{-}R_{nearest}^2$ for each variable. Small values of this ratio indicate good clustering.

International Clinical Research





RESULTS AND CONCLUSION

The cluster 1, considered the best separated, can be taken for normalisation purposes. Also the cluster 6 can still be considered well separated and useful for the normalisation purposes. Both the clusters contain some 'conservative' housekeeping genes. Alternatively, the clusters 1 and 6 can be joined. The encompassing cluster contains all the known housekeeping genes UBQ, Beta-actin, GAPD and 18S together with IGF-2, IGF-1R, BP-3 and BP-4. If a distinct cluster contains predominantly known housekeeping genes, its genes can be applied for normalization purposes in form of geometric mean as follows.

$$\text{Index} = \sqrt[n]{CP_1 \times CP_2 \times CP_3 \times \ldots\ldots \times CP_n}$$

where 1,2...n are the genes. Also genes, not *a priori* assumed to be unregulated, but those were tightly clustered with housekeeping genes can be included in the index.