ELSEVIER

Neuroscience Letters 339 (2003) 62–66

Neuroscience Letters

www.elsevier.com/locate/neulet

Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data

Christian Ramakers[a,b], Jan M. Ruijter[a,*], Ronald H. Lekanne Deprez[a], Antoon F.M. Moorman[a]

[a]*Department of Anatomy and Embryology K2-283, Experimental and Molecular Cardiology Group, Academic Medical Centre, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands*

[b]*Department of Cardiology, University Hospital Maastricht, Cardiovascular Research Institute Maastricht, Maastricht, The Netherlands*

Received 22 November 2002; received in revised form 4 December 2002; accepted 5 December 2002

Abstract

Quantification of mRNAs using real-time polymerase chain reaction (PCR) by monitoring the product formation with the fluorescent dye SYBR Green I is being extensively used in neurosciences, developmental biology, and medical diagnostics. Most PCR data analysis procedures assume that the PCR efficiency for the amplicon of interest is constant or even, in the case of the comparative C_t method, equal to 2. The latter method already leads to a 4-fold error when the PCR efficiencies vary over just a 0.04 range. PCR efficiencies of amplicons are usually calculated from standard curves based on either known RNA inputs or on dilution series of a reference cDNA sample. In this paper we show that the first approach can lead to PCR efficiencies that vary over a 0.2 range, whereas the second approach may be off by 0.26. Therefore, we propose linear regression on the Log(fluorescence) per cycle number data as an assumption-free method to calculate starting concentrations of mRNAs and PCR efficiencies for each sample. A computer program to perform this calculation is available on request (e-mail: bioinfo@amc.uva.nl; subject: LinRegPCR).

Keywords: mRNA quantification; Quantitative polymerase chain reaction (PCR); Real-time PCR; SYBR® green I; PCR efficiency; Linear regression; Standard curve method; Comparative C_t method

Recent developments in polymerase chain reaction (PCR) combined with new fluorescent techniques have led to the introduction of quantitative or real-time PCR [1,9]. This technique is extensively used in neurosciences, developmental biology and medical diagnostics and is preferred over other quantitative PCR methods because it does not rely on end-point analyzes, which can be misleading due to product inhibition, enzyme instability and a decrease of reaction components in time [2]. Quantification of cDNA in real-time PCR in combination with the double stranded DNA specific dye SYBR® Green I is based on the monitoring of the increasing fluorescence intensity after each PCR cycle. This SYBR green I monitoring is available in all quantitative PCR systems. All currently used data analysis methods are based on determining the threshold cycle (C_t), which is the fractional cycle number at which a fixed amount of DNA is formed. All calculations assume that the PCR efficiency of the amplicon of interest is constant over time and has the same value in all studied samples.

The comparative C_t method even assumes this constant efficiency to be equal to 2 for the target and the reference amplicon [6]. With this method, the amount of target amplicon X in sample S, normalized to an endogenous reference R and related to a control sample C is calculated as $2^{\{(-C_t^{X,S}-C_t^{R,S})-(C_t^{X,C}-C_t^{R,C})\}}$, resulting in the fold difference between sample and control. However, it has been reported that PCR efficiencies of target and reference amplicons can vary over a range of 1.8–2.0 [3]. When this variation in PCR efficiency is taken into account, it can easily be calculated that a real 10-fold difference can turn up as any value between 0.7 and 210 (Table 1). Even with a 0.04 range in PCR efficiencies (from 1.78 to 1.82) already a 4-fold error in fold-difference will occur. This shows that the comparative C_t method is very sensitive to variations in PCR efficiency. Alternatives, that include separate PCR efficiencies for target and reference amplicons but still assume

* Corresponding author. Tel.: +31-2-566-5386; fax: +31-2-697-6177.
E-mail addresses: j.m.ruijter@amc.uva.nl (J.M. Ruijter), c.r.ramakers@amc.uva.nl (C. Ramakers).

0304-3940/03/$ - see front matter © 2003 Elsevier Science Ireland Ltd. All rights reserved.
doi:10.1016/S0304-3940(02)01423-4

Table 1
Illustration of the effect of unequal PCR efficiencies on the result of the comparative C_t method[a]

	Sample		Control			Fold difference
	Amplicon X	Reference	Amplicon X	Reference		
N_0	0.1	0.01	0.01	0.01	→	Input: 10
E	2	1.8	1.8	2		
C_t	8.97	14.49	14.49	12.29	→	Obs: 212
E	1.9	1.9	1.9	1.9		
C_t	9.68	13.27	13.27	13.27	→	Obs: 12
E	1.8	2	2	1.8		
C_t	10.57	12.29	12.29	14.49	→	Obs: 0.7

[a] With this method, the fold difference defined as the amount of target amplicon X in sample S, normalized to an endogenous reference R and related to a control sample C can be calculated as $2^{\{(-C_t^{X,S}-C_t^{R,S})-(C_t^{X,C}-C_t^{R,C})\}}$. With an input 10-fold difference (N_0 line) and PCR efficiencies ranging from 1.8 to 2.0 (E lines) the fractional number of cycles needed to reach a threshold of 50 (N_t) can be calculated as $C_t = \{\mathrm{Log}(N_t) - \mathrm{Log}(N_0)\}/\mathrm{Log}(E)$. From these threshold cycles the observed (*Obs*) fold difference can be calculated [6].

constant efficiencies in all samples, have been published [3,7]. These alternatives rely on standard curves for the estimation of these PCR efficiencies.

The standard curve method is usually based on the C_t values of either an input series of known RNA concentrations in the reverse transcription or a dilution series of a reference cDNA sample [6]. The equation for PCR kinetics ($N_C = N_0 \times Eff^C$) states that the amount of product after C cycles (N_C) is equal to the starting concentration (N_0) times the efficiency (*Eff*) to the power C. This equation can be converted into the form $C_t = \mathrm{Log}(N_C)/\mathrm{Log}(Eff) - \{1/\mathrm{Log}(Eff)\} \times \mathrm{Log}(N_0)$ which is a linear relation between C_t and $\mathrm{Log}(N_0)$ when N_C is fixed and only when the efficiency is constant. The standard curve is constructed by fitting this equation to the $\mathrm{Log}(N_0)$ and C_t values for a known series of N_0. The PCR efficiency is then calculated from the slope of the standard curve ($Eff = 10^{-1/slope}$). However, in most applications of real-time PCR the standard curve-derived PCR efficiencies are not used. The standard curve is only used as a calibration curve to estimate the starting concentrations of unknown samples.

We scrutinized the use of standard curves for estimating PCR efficiencies by reviewing the two approaches for creating standard curves: (1) the use of (duplicate) serial dilutions of one of the cDNA samples under investigation; and (2) the use of a (duplicate) series of increasing amounts of total RNA as input in the reverse transcription [6]. When the standard curve is based on the serial dilution of a single cDNA sample the chance exists that the sample used to generate this standard curve is contaminated with for example salt, phenol, chloroform and/or ethanol resulting in a lower-than-expected PCR efficiency. However, dilution of this sample will also dilute the 'poison' and decrease its effect on the PCR reaction, thereby increasing the PCR efficiency with each dilution step. A computer calculation of this dilution effect (Fig. 1) clearly shows that there is no direct relation between the individual PCR efficiencies ($E_{diluted}$, 1.65–1.86) and the standard-curve-derived PCR efficiency (1.91). In this example the latter efficiency deviates 0.26 from the PCR efficiency of the undiluted reference sample and probably from all other samples prepared in the same way. Only when the individual PCR efficiencies of the standard curve samples are checked, a poisoned standard curve can be unmasked.

In the second approach, the standard curve is based on a series of increasing amounts of total RNA as input in the reverse transcription. To this end, six different concentrations of left ventricular total RNA of a mongrel dog were measured in triplicate. This experimental set-up enabled us to do a sub-sampling experiment and to construct two series of 3^6 (= 729) standard curves: one series based on six single data points per curve and another series based on six duplicate data points per curve. The frequency distributions of all slopes and calculated PCR efficiencies of these two series of standard curves are shown in Fig. 2. The calculated PCR efficiencies for standard curves based on single data points vary between 1.81 and 2.02; the use of duplicates halves this range (Fig. 2).

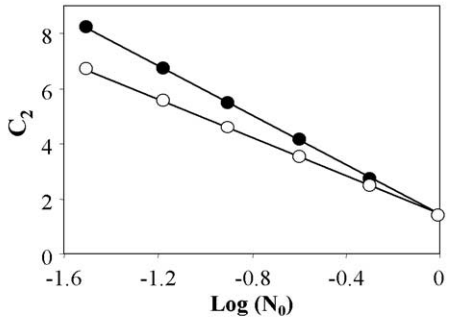
In summary, the comparative C_t method and the standard curve rely heavily on the assumption that in all samples the PCR efficiency of each amplicon is constant. To by-pass this assumption in the analysis of quantitative PCR data, the fluorescence measured per cycle of each sample can be used. This alternative approach is based on linearization of the basic formula for exponential PCR amplification by taking the logarithm on both sides of the equation resulting in $\mathrm{Log}(N_C) = \mathrm{Log}(N_0) + \mathrm{Log}(Eff) \times C$. The N_C and C are measured fluorescence data and cycle number, respectively. The log-linear part of the PCR data can be determined for each sample by selecting a lower and an upper limit of a 'window-of-linearity' (Fig. 3). Linear regression analysis is then used to calculate the intercept and the slope, Log (N_0)

Input

E_{max}	2.00	E_{lost} (= E_{max} − E_{actual})	0.35
E_{actual}	1.65		

Simulation

		poison partly dilutable (D = 0.6, U = 0.4)		poison not dilutable (D = 0, U = 1)	
Dilution	N_0	$E_{diluted}$	threshold cycle (C_2)	$E_{constant}$	threshold cycle (C_2)
1	1.0	1.65	1.38	1.65	1.38
2	0.50	1.755	2.47	1.65	2.77
4	0.25	1.808	3.51	1.65	4.15
8	0.125	1.834	4.57	1.65	5.54
16	0.062	1.847	5.54	1.65	6.79
32	0.031	1.853	6.74	1.65	8.31
∞		1.86			



Result

	Slope	Calculate Efficiency (= $10^{-1/slope}$)
●—	-4.60	1.65
○—	-3.55	1.91

Fig. 1. Computer simulation showing the influence of a poison in the cDNA sample that is serially diluted to construct a standard curve. **Input** gives the parameters used in the simulation. E_{max} represents the maximum PCR efficiency and E_{actual} the actual PCR efficiency due to a loss in PCR efficiency (E_{lost}). A fraction D of this loss is due to a poison that can be diluted, leaving a fraction U that cannot be diluted ($U = 1 - D$). **Simulation** shows the dilutions with the resulting N_0 values. The effect of the dilution on the PCR efficiency can be calculated with: $E = E_{max} - U \times E_{lost} - (D/dilution) \times E_{lost}$. This calculation is carried out for a situation in which part of the loss is due to a poison that can be diluted (column 3, $E_{diluted}$) and for a situation in which the loss in efficiency is due to factors, which cannot be diluted (column 5, $E_{constant}$). For both, the fractional number of cycles needed in each sample to reach the crossing line level of 2 ($C_2 = \{\text{Log}(2) - \text{Log}(N_0)\}/\text{Log}(E)$) is calculated in columns 4 and 6, respectively. The resulting threshold cycles (C_2) are plotted against the Log(N_0). Linear regression analysis of the 'poisoned' standard curve (open circles) shows perfect linearity ($R^2 = 1$). **Result** shows that the different slopes of the standard curves translate into significantly different calculated PCR efficiencies. Note that ultimate dilution of the poisoned sample would give an $E_{diluted}$ of 1.86.

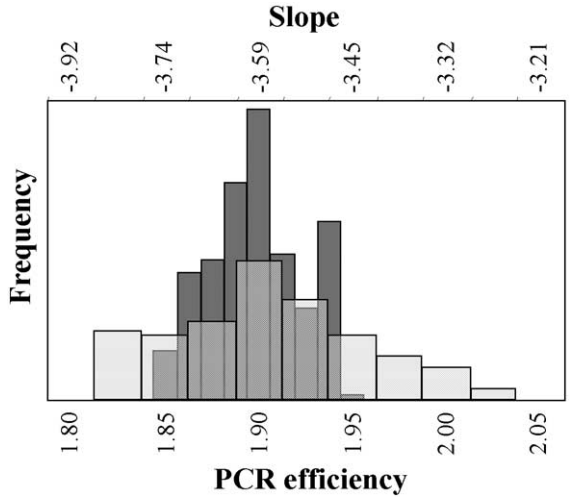


Fig. 2. Sub-sampling experiment to estimate the variability of the PCR efficiencies calculated from the slope of the standard curve. Six different concentrations of left ventricular total RNA of a mongrel dog (0.065, 0.13, 0.26, 0.65, 1.3, and 2.6 μg/50 μl) were reverse transcribed. In the real-time PCR, 1.3 μl of cDNA was used to amplify a 273 bp GAPDH amplicon using dog specific primers [8]. Details regarding reverse transcription and real-time PCR protocols have been previously published [4]. Based on samples in which reverse transcriptase enzyme was omitted during reverse transcription, genomic DNA contamination of the samples was estimated to be 15 000× less than the quantified cDNA levels and was therefore considered to be negligible. All cDNAs were quantified in triplicate. The graph shows the frequency distribution of PCR efficiencies that can be calculated from the slopes ($Eff = 10^{-1/slope}$) of 729 standard curves constructed from all possible permutations of the six times three RNA inputs (light bars). These PCR efficiencies range from 1.81 to 2.02. When the three possible duplicate observations per triplicate are used, the range of the distribution is reduced to 1.86–1.94 (dark bars). The correlation coefficient squared (R^2) in both sets of calibration curves ranged from 0.98 to 1.00 indicating almost perfect linearity of all curves.

and Log(Eff), respectively, from the straight line that fits best to the included data points. The starting concentration follows directly from the intercept of this linear regression line ($N_0 = 10^{intercept}$), and is expressed in terms of SYBR Green I fluorescence. The individual PCR efficiency follows from the slope of the linear regression line ($Eff = 10^{slope}$) and can be used as a quality check to exclude possible contaminated samples. To ensure unambiguous selection of data points within the window-of-linearity, an iterative algorithm is formulated to search for lines consisting of at least 4 and no more than six data points with the highest R^2 value and a slope close to the maximum slope. The above algorithm gives the best correlation between the 'known' input and the observed N_0 value as was tested in a large number of dilution series. Using lines with only three data points gives inconsistent N_0 values whereas longer lines may be biased by inclusion of below-detection-limit and/or plateau phase points. The slope criterion is included in the algorithm to avoid fitting in the plateau phase. Note that this algorithm provides an adequate 'window-of-linearity' for most if not all of the samples. However, visual inspection of all amplification curves remains necessary to identify deviating samples that require an individual window or may have to be excluded. A computer program that implements the linear regression analysis of real-time PCR data and the described algorithm, as well as the possibility to interactively draw a window-of-linearity, is available by automatic request (e-mail: bioinfo@amc.uva.nl; subject: LinRegPCR). This program is compatible with exported data

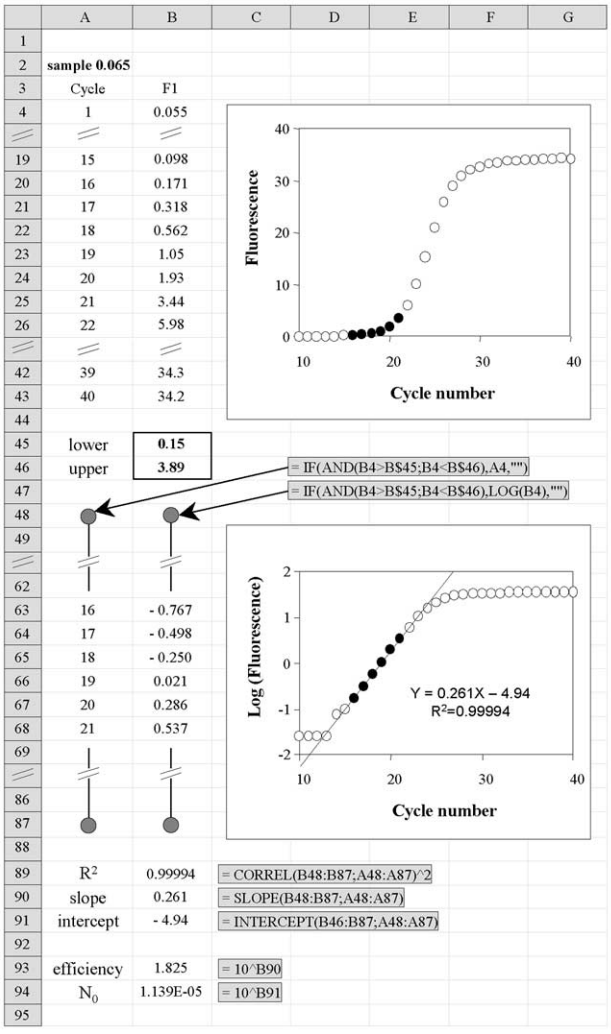


Fig. 3. Illustration of the linear regression calculations implemented in a Microsoft Excel® spreadsheet for determining starting concentrations and PCR efficiencies per sample. The cycle number – fluorescence data pairs are imported at starting position A2. Only data of one sample are shown with cycle number and fluorescence in columns A and B, rows 4 till 43. The IF-statements in rows 48 till 87 filter those data points that fall in the so-called window-of-linearity, determined by setting the values in the anchored cells B$45 and B$46. These values can be changed interactively. Once the window-of-linearity is established a regression line is fitted to the included data points. The functions in cells B89, B90 and B91 calculate the R^2, slope and the intercept, respectively. Finally, in cells B93 and B94 the PCR efficiency ($= 10^{slope}$) and N_0 ($= 10^{intercept}$) are calculated. A Microsoft Windows® program to perform these calculations is available by automatic request (e-mail: bioinfo@amc.uva.nl; subject: LinRegPCR).

from the LightCycler™ system (Roche) as well as the ABI Prism® systems (Applied Biosystems).

It has been proposed recently [5] to determine the PCR efficiency of each sample from the amount of fluorescence (N_A and N_B) and the number of cycles (C_A and C_B) at two arbitrary thresholds A and B along the exponential phase using the equation: $E = (N_A/N_B)^{1/(C_A - C_B)}$ (note that the symbols are adapted to the ones used in this letter and that a mistake in the exponent of the equation which appears in the original paper is corrected). There are several advantages of our proposed linear regression approach over this non-linear curve fit. The use of the *Log(fluorescence)* versus *cycle number* plot in the linear regression approach enables more reliable identification of below-detection-limit noise points and of data points that are no longer in the straight log-linear phase of the PCR reaction, both of which can not easily be identified in a non-logarithmic graph (Fig. 3: fluorescence versus Cycle number graphs). Furthermore, the use of all points in the log-linear phase with linear regression not only gives a more reliable measure for the slope of the regression line and thus for the PCR efficiency, but also provides the correlation coefficient as a measure for the goodness of fit. The main advantage of the linear regression approach is that an estimate of the N_0, the initial number of molecules, can directly be calculated from the intercept parameter of the regression line.

In conclusion, the 'poison' simulation as well as the sub-sampling experiment demonstrates that systematic and random differences in PCR efficiencies of individual samples make the standard curve an unreliable tool for quantification of mRNA levels. On the other hand, the linear regression approach enables the straightforward estimation and direct comparison of N_0 values between samples and between target and reference amplicons without the need to assume equal PCR efficiencies. This greatly simplifies the calculation involved in real time PCR analyzes and enables an independent estimation of mRNA contents in single samples. This may be of special importance when low quantities have to be reliably measured in small tissue samples as in fields of neuroscience and medical diagnostics.

Acknowledgements

We would like to thank Dr Marc Vos, Department of Cardiology, Academic Hospital Maastricht, and Dr Onno Bakker, Department of Endocrinology and Metabolism, Academic Medical Center, Amsterdam for their constructive support. This research is partly supported by the Netherlands Heart Foundation, grant NHS 98.042 (CR) and NHS 96.002 (RHLD) (AFMM).

References

[1] U.E.M. Gibson, C.A. Heid, P.M. Williams, A novel method for real time quantitative RT-PCR, Genome Res. 6 (1996) 995–1001.
[2] P. Kains, The PCR plateau phase – towards an understanding of its limitations, Biochim. Biophys. Acta 1494 (2000) 23–27.
[3] W. Kamphuis, A. Schneemann, L.M. Van Beek, A.B. Smit, P.F.J. Hoyng, E. Koya, Prostanoid receptor gene expression profile in human trabecular meshwork: a quantitative real-time PCR approach, Invest. Opthalmol. Vis. Sci. 42 (2001) 3209–3215.
[4] R.H. Lekanne Deprez, A.C. Fijnvandraat, J.M. Ruijter, A.F.M. Moorman, Sensitivity and accuracy of quantitative real-time PCR

using SYBR green I depend on cDNA synthesis conditions, Anal. Biochem. 307 (2002) 63–69.
[5] W. Liu, D.A. Saint, A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics, Anal. Biochem. 302 (2002) 52–59.
[6] Livak, K.J., ABI Prism 7700 Sequence Detection System. User bulletin 2, PE Applied Biosystems (2001).
[7] J. Meijerink, C. Mandigers, L. van de Locht, E. Goodsaid, J. Raemakers, A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR, J. Mol. Diagn. 3 (2001) 55–61.
[8] S.C. Verduyn, C. Ramakers, G. Snoep, J.D.M. Leunissen, H.J.J. Wellens, M.A. Vos, Time course of structural adaptations in chronic AV block dogs: evidence for differential ventricular remodeling, Am. J. Physiol. Heart Circ. Physiol. 280 (2001) H2882–H2890.
[9] C.T. Wittwer, K.M. Ririe, R.V. Andrew, D.A. David, R.A. Gundry, U.J. Balis, The LightCycler™: a microvolume multisample fluorimeter with rapid temperature control, Biotechniques 22 (1997) 176–181.