The Importance of and a Method for Including Transfection Efficiency Into Real-Time PCR Data Analyses

WT. Godbey, Xiujuan Zhang, Fengqi Chang
Laboratory for Gene Therapy and Cellular Engineering, Department of Chemical & Biomolecular Engineering, Tulane University, 6823 St. Charles Ave., 300 Lindy Boggs Center, New Orleans, Louisiana 70118; telephone: 504-865-5872; fax: 504-865-6744; e-mail: godbey@tulane.edu

Received 5 October 2007; revision received 7 December 2007; accepted 2 January 2008
Published online 28 January 2008 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.21811

ABSTRACT: The polymerase chain reaction (PCR) is widely used to ascertain absolute or relative changes in the expression levels of specific genes as a function of cell type or in response to changes in environmental stimuli. Real-time PCR is an advance which allows for the analysis of gene expression over a wide range of initial cDNA concentrations, where the cDNA is the product of reverse transcriptase reactions applied to RNA samples. With the advent and advances in gene delivery technologies, it is now common for the cellular responses under scrutiny to be initiated via the expression of an exogenously delivered gene. When transfection (or transduction) is a part of the procedure used to prepare cell samples for real-time PCR, it is necessary to take the efficiency of gene delivery into account. Here a robust mathematical model for such analyses is derived, and validated with theoretical and experimental support. Comparison to existing analysis methods is presented to demonstrate the high significance of noting transfection, loading, and primer PCR efficiencies when processing PCR data. Biotechnol. Bioeng. 2008;XXX: XXX–XXX.

KEYWORDS: real-time PCR; gene delivery; transfection efficiency; Delta Delta; polymerase chain reaction

Introduction

The polymerase chain reaction (PCR), an enzymatic method that amplifies specific DNA sequences based on carefully designed primer sets, was originally used to produce usable copies of specific genes from nanogram amounts of DNA (Scharf et al., 1986). Typically, whole-cell RNA is collected from samples and cDNA is generated prior to amplification with DNA polymerase. Amplifications take place via a thermocycler, and results can be analyzed in a variety of ways, including agarose electrophoresis or hybridization. Traditional PCR analyses are labor-intensive and involve post-PCR sample processing, which add time, cost, and potential error to the data acquisition process. An improvement that utilized the 5′–3′ exonuclease activity of Thermus aquaticus (Taq) (Holland et al., 1991) led to the advent of real-time PCR (Heid et al., 1996). Real-time PCR incorporates dual-labeled probes that hybridize to amplicons during the PCR process. The probe labels consist of a fluorophore and a quencher, where the quencher absorbs photons with wavelengths in the emission spectra of the fluorophore (Lee et al., 1993; Morrison et al., 1989). Nuclease degradation of the hybridized probe releases fluorophore that can be spectrally detected. As amplification continues, detectable fluorophore amounts increase in a predictable fashion with each PCR cycle. With a fluorescence detection event occurring with each round of PCR amplification, results are obtained in a near real-time fashion—hence the name. The technology has since advanced to allow for the detection of four cDNA sequences simultaneously (Vet et al., 1999).

Instead of using the above mentioned TaqMan probe to detect the amplicon quantity, one could elect to use a non-sequence specific double-stranded DNA (dsDNA) binding dye such as SYBR Green I. The fluorescence of this dye, when bound to dsDNA, is more than 1,000-fold higher than that of the free dye (Morrison et al., 1998; Wittwer et al., 1997). Thus, amplification of a given gene can be monitored in real time via increases in fluorescence that indicate an increase in the amount of dsDNA produced by DNA polymerase. Although SYBR Green I is prone to lead to false positive assays due to its non-sequence-specific binding, it can work very well in single-product PCR reactions with appropriately designed primers that prevent non-specific background accumulation (e.g., primer-dimers) in very late cycles (Pfaffle, 2004; Wilhelm and Pingoud, 2003). SYBR Green I is
convenient, relatively inexpensive, and flexible in that it can be used in separate assays for different cDNAs. However, multiplex reactions with this dye are not possible.

Real-time PCR can be used to indicate qualitative or quantitative (Heid et al., 1996) mRNA amounts. Perhaps the two most common methods used for qualitative transcriptional assessment are the $2^{-\Delta C_T}$ (also called the Delta Delta $C_T$ method) (Livak and Schmittgen, 2001) and the Efficiency-corrected $C_T$ method (Pfaffl, 2001). The $2^{-\Delta C_T}$ method takes loading efficiencies (pipettor error) into account by normalizing machine outputs to detected levels of reference genes for which cellular levels remain approximately constant. Examples of reference genes include genes coding for ribosomal subunits (i.e., 18S and 28S rRNA), β-2 microglobulin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and TATA box-binding protein (Li et al., 2005; Schmittgen and Zakrajsek, 2000). The Efficiency-corrected $C_T$ method goes a step further by incorporating primer PCR efficiencies for primer pairs into output analyses. [The original $2^{-\Delta C_T}$ paper also does this, but drops the efficiencies for simplicity (Livak and Schmittgen, 2001).] Both calculation methods are straightforward and quite useful, but neither is suited for cases where gene delivery has been used as part of the experimental setup.

When real-time PCR experiments are performed, results are obtained for an entire sample of cells from which RNA has been obtained. When the experiments are designed to identify or measure cellular responses to the products of exogenously delivered genes, such as when a gene for a signaling molecule is transfected into cells and the resulting up- (or down-) regulation of genes downstream in a cascade are to be noted, then the results initially obtained from real-time PCR will reflect the transcriptional status of genes in both transfected and untransfected cells if the transcription efficiency is anything below 100%. However, if the transfection efficiency is known, then the cellular responses of only the subset of positively transfected cells within the sample can be determined. Here we present a mathematical method for deducing the relative change in transcriptional levels of genes in transfected cells by utilizing PCR information that is corrected for loading, primer, and transfection efficiencies.

This model could be useful for the determination of genetic pathways or cellular responses, at the genetic level, to specific proteins. While some groups administer exogenous signals via cell media to study cellular responses, gene delivery opens the possibility of using a broader set of internal signals. For instance, by delivering a gene that codes for a transcription factor, one could analyze the transcriptional up- or down-regulation of a related set of genes via real-time PCR without relying upon gene array technology. This technique could also act as a supplement to Western blot analyses, which are used to elucidate protein expression levels (which can vary from transcriptional levels of the corresponding gene). Artificial cascades could also be developed by cellular engineers through the delivery of one or more genes that are not normally expressed by that cell, or the relation of protein (over)expression to oncogenesis could be implicated through transfection and real-time PCR. Real-time PCR has been used for years, but the model described herein allows for the coupling of this technique with gene delivery.

**Materials and Methods**

**Cell Culture and Transfection**

Mouse fibroblast cells (NIH/3T3) were grown in Dulbecco's modified Eagle's medium (DMEM; Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum and maintained in a humidified incubator at 37°C and 5% CO₂. Cells were plated at a density of approximately 104 cells/mm² 18 h prior to transfection. Cells were transfected with 3.6 µg DNA/10⁵ cells, using poly(ethyleneimine) (PEI) with weight-average MW = 25 kDa (Sigma-Aldrich, St. Louis, MO) as the gene delivery vehicle (Godbey et al., 1999). PEI/DNA complexes were constructed using a 7.5:1 PEI nitrogen to DNA phosphate ratio.

**Real-Time Reverse Transcription-PCR Analysis**

NIH/3T3 cells were transfected as described above using the pEGFP-N1 or pDsRed vector (Clontech (Takara Bio Company), Mountain View, CA). The transfected cells were observed under an Olympus IX71 fluorescence microscope (Olympus America, Center Valley, PA) and pictures were taken 24 h after transfection. Each transfected sample (including transfected cells and non-transfected cells), was divided into two tubes, and each tube was subjected to fluorescence activated cell sorting (FACS) (FACS Vantage/ DiVa, Becton Dickinson, San Jose, CA). For one fraction, the FACS apparatus was set with a threshold value of zero so that all cells were gated for later analysis, and the other fraction had FACS performed with a positive threshold that only gated GFP-positive cells. Total RNA was extracted from each cell population (and a negative control consisting of untransfected cells) immediately after sorting using an RNeasy kit (Qiagen, Valencia, CA), and isolated RNA was treated with DNase I to remove any traces of genomic DNA. First-strand cDNA was synthesized from 1 µg of each RNA sample using random hexamers and reverse transcriptase (Bio-Rad, Hercules, CA). Real-time reverse transcription-PCR was performed with Sybr Green using EGFP-specific primers and primers for the reference gene 18S rRNA (Table I) in an iCycler iQ system (Bio-Rad). Cycling parameters were 95°C for 5 min to activate Taq DNA polymerase, then 40 cycles of 95°C for 15 s, 58°C for 30 s, and 72°C for 20 s. To confirm the specificity of PCR products, melting curves were determined using iCycler software and samples were run on a 2% agarose gel. 18S rRNA was used as the internal control.
Primer PCR efficiencies were calculated by running regression lines through data points on plots of \( C_t \) values versus the logarithms of various known starting quantities of cDNA. Starting quantities of cDNA were obtained via serial dilutions over 3–5 orders of magnitude for each primer set and were performed in triplicate. \( C_t \) values were determined by iCycler iQ software, version 3.1 (Bio-Rad). Regression lines had correlation coefficients \((R^2)\) values in excess of 0.97. The amplification efficiency \( \eta \) was defined as \( 10^{(-1/slope)} \), where slope = slope of the regression line for the given primer set. (BioRad, 2006) The primer PCR efficiency was then calculated as \( E = (\eta - 1) \).

### Table 1.

<table>
<thead>
<tr>
<th>Primer set</th>
<th>Primer sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFP primer set</td>
<td>For: 5'-ATC ATG GCC GAC AAG CAG AAG AAC-3'</td>
</tr>
<tr>
<td></td>
<td>Reverse: 5'-GTA CAG CTC GTC CAT GCC GAG AGT-3'</td>
</tr>
<tr>
<td>18S rRNA primer set</td>
<td>For: 5'-CGG ACA CGG ACA GGA TTG ACA GAT-3'</td>
</tr>
<tr>
<td></td>
<td>Reverse: 5'-TCA ATC TCG GGT GGC TGA ACG C-3'</td>
</tr>
</tbody>
</table>

### Statistics

Pairwise comparisons were carried out via heteroscedastic \( t \)-tests, \( n \geq 3 \). Groups of data were analyzed via single-factor ANOVA.

### Theoretical Aspects

For a given real-time PCR experiment, there will be a negative control run (neg) and \( n \) sample runs (sample\_\( i \), \( 1 \leq i \leq n \)). Within a given run there will be an internal control (reference gene \( ref \)) and \( m \) genes under investigation (gene\_\( i \), \( 1 \leq i \leq m \)).

In addition, when a set of cells undergoes gene delivery, if the delivery efficiency is less than 100\% there will be transfected and untransfected cells in the sample. The following notations are also used:

- \( tfct \) = a value associated with only positively transfected cells (assuming no net effect for untransfected cells)
- \( untransfected \) = a value associated with only untransfected cells
- \( obs \) = the value observed for an entire sample that contains both transfected and untransfected cells.
- (N.B.: \( [RNA]_{gene,sample} \) represents the RNA concentration from transfected cells, normalized to depict a sample that is completely transfected.)

Although real-time PCR does not directly return RNA concentrations for given samples, the entire experimental method is based upon initial RNA concentrations. For a given transfected sample, the initial RNA concentration used to make cDNA for the polymerase chain reaction is made up of contributions from both transfected and untransfected cells. Letting “\( \% \)” represent the transfection efficiency (the percentage of transfected cells, assuming equal expression of the delivered gene in all positively transfected cells),

\[
\text{obs}[RNA]_{gene,sample} = x\%\text{tfct}[RNA]_{gene,sample} - (1 - \%\text{untransfected})[RNA]_{gene,sample}
\]

where \( x = ([RNA]_{ref,sample}/[RNA]_{ref,neg}) = a \) scaling factor to take into account loading errors.

Assuming that untransfected cells express gene\_\( i \) in the same manner as that of negative controls, we can replace \( \text{untransfected}[RNA]_{gene,sample} \) with \( [RNA]_{gene,neg} \) and rearrange to get the following expression for the RNA concentration in positively transfected cells:

\[
\text{tfct}[RNA]_{gene,sample} = ((\text{obs}[RNA]_{gene,sample}/x) - (1 - \%)[RNA]_{gene,neg})
\]

Real-time PCR relies on the Cycle Threshold (\( C_t \)) number, defined as the number of cycles (ideally, doublings) required to produce a minimally detectable amount of DNA copies (fluorescence). The detection threshold can be considered arbitrary, and it remains constant over all runs in an experimental set.

Define \( k \equiv \) the threshold value, over which concentrations of DNA are reliably detectable.

As mentioned above, in the ideal case the amount of DNA is doubled with every PCR cycle. However, it is usually the case that the primer PCR efficiency is not 100\%. Let \( E_{gene_i} \) (or \( E_{ref} \)) \equiv \) the primer PCR efficiency of the primer set for a given gene. Note that this efficiency is constant for a given gene over all runs (samples \( 1 - j \) and the negative control).

Then, in general, \((E + 1)^{C_t}[RNA] = k\)

\[
\Rightarrow (E + 1)^{C_t} = \frac{k}{[RNA]}
\]

\[
\Rightarrow C_t \ln(E + 1) = \ln\left(\frac{k}{[RNA]}\right)
\]

\[
\Rightarrow C_t = \frac{\ln(k/[RNA])}{\ln(E + 1)}.
\]
Combining this relation with (Equation 1) yields the following relation for the $C_t$ value that would be obtained if the gene delivery were carried out at 100% efficiency:

$$tfeC_{gene, sample}^i = \frac{\ln(k / ((obs[RNA]_{gene, sample} / x) - (1 - \%)[RNA]_{gene, neg} / \%)) \ln(E_{gene} + 1)}{\ln(k / ((obs[RNA]_{gene, sample} / x) - (1 - \%)[RNA]_{gene, neg} / \%))}$$

$$= \frac{\ln((k / obs[RNA]_{gene, sample}) (\%/1/x) - ((1 - \%)[RNA]_{gene, neg} / obs[RNA]_{gene, sample})) \ln(E_{gene} + 1)}{\ln(E_{gene} + 1)}$$

$$= \frac{\ln((k / obs[RNA]_{gene, sample}) + \ln(\%/1/x) - ((1 - \%)[RNA]_{gene, neg} / obs[RNA]_{gene, sample})) \ln(E_{gene} + 1)}{\ln(E_{gene} + 1)}$$

$$\Rightarrow tfeC_{gene, sample}^i = obs C_{gene, sample}^i + \frac{\ln(\%/1/x) - (1 - \%)([RNA]_{gene, neg} / obs[RNA]_{gene, sample})) \ln(E_{gene} + 1)}{\ln(E_{gene} + 1)}$$

While valid, this equation is not directly useful without knowing the original concentrations of RNA for the gene of interest in the negative control and the test sample. While it is possible to calculate these values, such calculations are unnecessary since they can be replaced by functions of the corresponding $C_t$ values:

$$\Rightarrow obs C_{gene, sample}^i = \frac{\ln(k / obs[RNA]_{gene, sample}) \ln(E_{gene} + 1)}{\ln(E_{gene} + 1)}; C_{neg, gene}^i = \frac{\ln(k / [RNA]_{gene, neg}) \ln(E_{gene} + 1)}{\ln(E_{gene} + 1)}$$

$$\Rightarrow obs C_{gene, sample}^i \ln(E_{gene} + 1) = \ln\left(\frac{k}{obs[RNA]_{gene, sample}}\right)$$

$$\Rightarrow (E_{gene} + 1)^{obs C_{gene, sample}^i} = \frac{k}{obs[RNA]_{gene, sample}}; (E_{gene} + 1)^{C_{neg, gene}^i} = \frac{k}{[RNA]_{gene, neg}}$$

$$\Rightarrow \frac{(k / obs[RNA]_{gene, sample})}{(k / [RNA]_{gene, neg})} = \frac{(E_{gene} + 1)^{obs C_{gene, sample}^i}}{(E_{gene} + 1)^{C_{neg, gene}^i}}$$

$$\Rightarrow \frac{[RNA]_{gene, neg}}{obs[RNA]_{gene, sample}} = (E_{gene} + 1)^{obs C_{gene, sample}^i} - C_{neg, gene}^i$$

Substituting (Equation 3) into (Equation 2) yields an expression for $C_t$ that takes transfection efficiency into account:

$$tfeC_{gene, sample}^i = obs C_{gene, sample}^i$$

$$+ \frac{\ln(\%/1/x) - (1 - \%)\ln(E_{gene} + 1)obs C_{gene, sample}^i - C_{gene, neg}^i \ln(E_{gene} + 1)}{\ln(E_{gene} + 1)}$$

Returning to $x$, the scaling factor used to take into account loading errors, we can plug in a rearrangement of the previously derived formula for $C_t$ (see above) to remove direct reliance upon RNA concentrations:

$$(E + 1)^{C_t} = \frac{k}{[RNA]}$$

$$\Rightarrow [RNA] = \frac{k}{(E + 1)^{C_t}}$$

Then,

$$x = \frac{[RNA]_{ref, sample}^i}{[RNA]_{ref, neg}^i} = \frac{(k / (E_{ref} + 1)^{C_{ref, sample}^i})}{(k / (E_{ref} + 1)^{C_{ref, neg}^i})}$$

$$= \frac{(E_{ref} + 1)^{C_{ref, neg}^i}}{(E_{ref} + 1)^{C_{ref, sample}^i}} = (E_{ref} + 1)^{C_{ref, sample}^i} - C_{ref, sample}^i$$

$$\Rightarrow \frac{1}{x} = (E_{ref} + 1)^{C_{ref, sample}^i} - C_{ref, sample}^i$$
Plugging (Equation 5) into (Equation 4) yields the final expression for the normalized \( C_\text{t} \) value for positively transected cells in a sample, taking into account transfection efficiency, primer PCR efficiency, and loading error:

\[
\text{tct} \, C_{\text{gene, sample}} = \text{obs} \, C_{\text{gene, sample}} + \ln \left( \frac{\%}{(E_{\text{gene}} + 1) \, C_{\text{gene, sample}}^{-1} - C_{\text{gene, sample}}^{-1} - (1 - \%) \, (E_{\text{gene}} + 1) \, \text{obs} \, C_{\text{gene, sample}}^{-1} - C_{\text{gene, neg}}^{-1}}{\ln(E_{\text{gene}} + 1)} \right)
\]

Now that a true, normalized \( C_\text{t} \) value has been obtained, it can be placed into the following expression to yield the factor by which transfection changes RNA levels:

\[
\Rightarrow \, \text{factor} = (E_{\text{gene}} + 1) \, C_{\text{gene, neg}}^{-1} - (\text{obs} \, C_{\text{gene, sample}}^{-1} + \ln \left( \frac{\%}{(E_{\text{gene}} + 1) \, C_{\text{gene, sample}}^{-1} - C_{\text{gene, sample}}^{-1} - (1 - \%) \, (E_{\text{gene}} + 1) \, \text{obs} \, C_{\text{gene, sample}}^{-1} - C_{\text{gene, neg}}^{-1}}{\ln(E_{\text{gene}} + 1)} \right))
\]

\[
= (E_{\text{gene}} + 1) \, C_{\text{gene, neg}}^{-1} - (\text{obs} \, C_{\text{gene, sample}}^{-1} + \ln \left( \frac{\%}{(E_{\text{gene}} + 1) \, C_{\text{gene, sample}}^{-1} - C_{\text{gene, sample}}^{-1} - (1 - \%) \, (E_{\text{gene}} + 1) \, \text{obs} \, C_{\text{gene, sample}}^{-1} - C_{\text{gene, neg}}^{-1}}{\ln(E_{\text{gene}} + 1)} \right))
\]

\[
= (E_{\text{gene}} + 1) \, C_{\text{gene, neg}}^{-1} - \left((E_{\text{gene}} + 1) \, C_{\text{gene, sample}}^{-1} - C_{\text{gene, sample}}^{-1} - (1 - \%) \, (E_{\text{gene}} + 1) \, \text{obs} \, C_{\text{gene, sample}}^{-1} - C_{\text{gene, neg}}^{-1} \right)
\]

\[
= (E_{\text{gene}} + 1) \, C_{\text{gene, neg}}^{-1} - (1 - \%)
\]

Note that when transfection efficiency is equal to 100%, this equation reduces to the Efficiency-corrected \( C_\text{t} \) model. It should also be noted that, if the PCR efficiencies of each primer pair are assumed to be 1.0, then (Equation 7) can be reduced to:

\[
2^{-\Delta C_\text{t}} - (1 - \%)
\]

where \( 2^{-\Delta C_\text{t}} \) is the same value used by many and first described by Livak and Schmittgen (2001). However, users should be cautioned that analyzing real-time PCR data without knowledge of the efficiencies of each PCR primer pair could produce erroneous results. For instance, if two primer pairs (such as for the reference gene and the gene of interest) had PCR efficiencies of 1.0 and 0.98, respectively, then the initial RNA amount of the gene of interest would be miscalculated by a factor of \( 1.9 \times 10^5 \) after 20 cycles if the primer PCR efficiency of the primer set for the gene of interest were rounded up to 1.0. Since acceptable primer efficiencies are generally accepted to be between 0.90 and 1.05, (BioRad, 2006) the amount of error associated with ignoring primer PCR efficiency could become significant.

### Results and Discussion

#### Rudimentary Example

To illustrate the importance of this model, consider the following example (Fig. 1). Suppose we have a gene that is expressed such that there are typically 8 copies of mRNA per cell, and that we are able to produce cDNA such that there will be 1 cDNA molecule for every mRNA in a cell. For such a case, when we observe the relative mRNA concentration for that gene, it would be equal to 8 (arbitrary units). Now consider a stimulus that could amplify the expression of the gene in question by a factor of two. If the stimulus came about as the result of the expression of a delivered gene, the transfection efficiency would play a key role in the amount of gene stimulation we would be able to observe. At 25% transfection efficiency, the observed RNA concentration would be equal to 10 arbitrary units: 25% of the cells would display the \( 2 \times \) amplification, and 75% of the cells would express the gene at a level assumed to be the same as that of the negative control. Now suppose that the threshold value, \( k \), is set to 32, and the primer PCR efficiency of all primer sets is 100%. Also assume perfect loading, so that \( C_{\text{ref, sample}} = C_{\text{ref, neg}} \). It would take two cycles (two doublings) of the negative control (8) to get our cDNA concentration to the detectable threshold value. Expressed differently,

\[
C_{\text{gene, neg}} = \frac{\ln((k/|RNA|))}{\ln(E + 1)} = \frac{\ln(4)}{\ln(2)} = 2
\]

Or, strictly in terms of doubling,

\[
C_{\text{neg}} = \log_2 \left( \frac{k}{|RNA|_{\text{neg}}} \right) = \log_2 \left( \frac{32}{8} \right) = 2
\]
The \( C_t \) value returned by the analysis software for the transfected sample would be equal to

\[
\ln(32/10) \quad \ln(1 + 1) = 1.678
\]

but if expression of the gene in question were amplified by \( 2x \), the \( C_t \) number should be one less than the negative control because one less doubling would be needed to amplify the cDNA to the threshold value. Utilizing (Equation 6) and the two \( C_t \) values returned by the analysis software, we get:

\[
t_{\text{ref}} C_{\text{gene, sample}} = 1.678 + \frac{0.25}{\ln(1 + 1)} = 1
\]

and the factor of change for the gene in question as a result of expression of the delivered gene is \( 2^{1.678} = 2 \). Both of these values can be verified by hand.

**Theoretical Comparison to Other Analysis Methods**

For an experiment that is designed to assess changes in transcription of a specific gene or genes in response to a transfection (or transduction) event, any analyses that involve \( C_t \) will have results that are dependent upon transfection efficiency. Two methods have been published and widely used for traditional real-time PCR results assessment. The first, as has already been mentioned, is often referred to as the \( 2^{-\Delta C_t} \) method (Livak and Schmittgen, 2001). This approach rounds the PCR efficiencies of all primer pairs to 2, and plugs machine-generated \( C_t \) data into the equation

\[
2^{-(C_{\text{target}} - C_{\text{ref}})_{\text{sample}}} = 2^{-(C_{\text{target}} - C_{\text{ref}})_{\text{sample}} - (C_{\text{target}} - C_{\text{ref}})_{\text{ref}}}
\]

The other method, published by Pfaffl (2001) and referred to as the Efficiency-corrected \( C_t \) model (Pfaffl, 2001) astutely uses values \( E \) that are equal to primer PCR efficiencies +1 to return relative expression ratios.

**Relative Expression Ratio**

\[
\text{Relative Expression Ratio} = \frac{(E_{\text{target}} - C_{\text{neg}})}{(E_{\text{reference}} - C_{\text{neg}})}
\]

where \( E \equiv 10^{(-1/\text{slope of one cycle in exponential phase})} \) Setting \( E \) values equal to 2 (primer efficiencies = 1.0 + 1) and rearranging terms yields an expression identical to the \( 2^{-\Delta C_t} \) equation.

Consider the following hypothetical example, where gene \( x \) has been delivered to cells to investigate the effects of \( x \) upon the transcription of a target gene:

This example demonstrates that the consideration of the

**Experimental Support**

To test the validity of the transfection efficiency model, cells were transfected with a gene coding for an enhanced green
fluorescent protein (GFP) and real-time PCR was performed to detect relative levels of GFP mRNA. Two classes of samples were used: untreated controls and transfected cells. The transfected cells were subsequently divided into two groups: one group that was run through a cell sorter with 100% of the cells being gated for further analysis, and another group that was separated into a (near) pure GFP-expressing fraction via FACS. Transfection efficiencies were determined for each group via FACS. As expected, the observed $C_t$ values of the transfected versus the transfected-and-sorted cell populations were different, but when the new model was applied to the heterogeneous cell population, a factor of transcriptional change was calculated that was within a statistically relevant range to near-pure GFP-expressing samples (Table II). In cases where no cDNA was detected for a specific gene, such as for the negative control in this experiment, a $C_t$ value equal to 1 + the number of PCR cycles performed ($C_t = 41$ in this case) was used for calculations.

When the data were analyzed with the $2^{-ΔΔC_t}$ and Efficiency-corrected $C_t$ models, the returned expression factors for the mixed and sorted cell populations were significantly different. However, applying the transfection efficiency model to the data yielded roughly equivalent $C_t$ values and expression factors for the two groups, demonstrating the validity of the model (Fig. 2). The data also illustrate the limitations of the $2^{-ΔΔC_t}$ and Efficiency-corrected models for analyzing populations of cells where transfection has been used to bring about a cellular response. These two models are not designed to tease out data from mixed populations of cells, and Figure 2 shows that there is significant statistical error associated with making claims regarding the effects of one gene product upon the expression of other genes when the signaling gene is not expressed in all of the cells in the population.

Table II. Observed $C_t$ values for the experiment depicted in Figure 2 (solid black, solid white bars only).

<table>
<thead>
<tr>
<th></th>
<th>$C_t$obs</th>
<th>18s</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative control</td>
<td>9.3</td>
<td>9.1</td>
</tr>
<tr>
<td>Mixed cell population</td>
<td>14.0</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>13.8</td>
<td>9.0</td>
</tr>
<tr>
<td>Sorted cell population</td>
<td>12.3</td>
<td>8.6</td>
</tr>
</tbody>
</table>

The transfection efficiency for the mixed cell population was 42.2%. The percentage of GFP-positive cells in the sorted population was 97.3%. Primer PCR efficiencies: GFP = 1.032, 18s RNA = 0.972.

For the above experiment, a population of untransfected cells was also sorted out of the mixed population and analyzed via real-time PCR in an attempt to verify the assumption that untransfected cells behaved like negative control cells, as far as transcription of the gene of interest was concerned. Due to technical limitations of cell sorting, it was infeasible to obtain a pure sample of untransfected cells—a sample of 94.1% untransfected cells was obtained. When this sample was further analyzed, though, a very interesting result was produced. As expected, the transfected cell “contamination” was detected by real-time PCR. However, when the sample was relabeled as a mixed population and plugged into (Equation 7), using 5.9% (1.0–94.1%) as the transfection efficiency, a value very similar to the value obtained for the sorted population was returned, adding credence to the model presented here (Fig. 2, striped bars).

An additional experiment was run to test the assumption that $C_t$ values for untransfected cells could be replaced by the $C_t$ values of the negative controls for the GFP transfection trial. This time, the plasmid DsRed was delivered into cells and transfection was verified visually to be approximately 30%. When the entire sample was subjected to real-time PCR to detect GFP expression, no GFP transcription was detected (data not shown), indicating that, in terms of pEGFP delivery, untransfected cells can viewed as behaving like negative controls. (“Untransfected,” in this case, means cells that do not transcribe and translate the GFP gene.)

Potential Problem

While use of the (Equation 7) is meant to return the factor by which the level of RNA for a given gene or genes is
changed following gene delivery (or gene delivery and subsequent cellular stimulation), we have noted the return of meaningless values in certain cases that involve down regulation. When the quantity

\[
1 - \left\{ \left( \frac{E_{\text{gene}, neg}}{E_{\text{ref}} + 1} \right)^{\frac{C_{\text{gene}, neg}}{C_{\text{gene}, sample}}} - \left( \frac{C_{\text{gene}, sample}}{C_{\text{gene}, neg}} \right) \right\}
\]

(or \(1 - 2^{-\Delta C_t}\), for short) > %, then a negative number will be returned by the equation. This scenario is theoretically impossible. Returned factors between zero and one are used to indicate the amount of down regulation of a given gene, but a negative number has no physical interpretation. In our experience, negative numbers have been returned in cases where the average factor, plus or minus one standard deviation, includes the value zero. In such a range, sensitivity is a key issue. In our worst case scenario (data not shown), a factor of \(-4.01\) was returned, but an alteration of the data by 4.2% was enough to yield a positive factor. The point here is that uninterpretable negative factors can be reasonably explained by machine, and/or pipettor error, each of which can be reasonably expected to be in the 3–5% range.

Another possible explanation for the return of negative numbers by (Equation 7) would be when the assumptions do not hold. The key assumption to be aware of is: an untransfected cell in the sample will behave identically to cells in the negative control, or, mathematically, \(\text{untransfected}_{\text{RNA}}^{\text{gene, sample}} = \text{RNA}_{\text{gene, neg}}\). However, if a transfected cell were to produce an excreted molecule that signaled neighboring cells to alter their transcription of the gene in question, then it is possible for the equation to yield a negative value. The appearance of a negative value should serve as an alert to the investigator that complex signaling may be taking place. Such an alert is not given by any other real-time PCR analysis method.

This model of analysis of real-time PCR data is very important when gene delivery has been used to bring about a given response. While it seems that cell sorting could be used to get around the issue of transfection efficiencies below 100%, not every laboratory has access to a cell sorter. In addition, sorting out 100% pure populations is technically infeasible. With this model in place, researchers now have a strengthened tool for analyzing gene expression cascades and cellular responses to specific intracellular proteins.

Conclusions

Real-time PCR is a valuable tool for assessing relative or actual RNA concentrations in cell populations. However, if the effects under investigation are being brought about via gene delivery, then the transfection efficiency for each sample is of great importance and should be incorporated into all calculations that involve the \(C_t\) value. This is especially true for non-viral gene delivery, which is known to yield transfection efficiencies well below 100%. Here we have presented a mathematical method of analysis that is more robust than existing real-time PCR analytical methods when gene delivery is an integral part of the investigation.

The authors wish to thank Prof. John Prindle for his valuable insights and discussions.

References


