Nucleic Acids Research Advance Access published June 25, 2012

*Nucleic Acids Research, 2012, 1–9*
*doi:10.1093/nar/gks588*

# Quantification of RNA integrity and its use for measurement of transcript number

**Michael J. Brisco and Alexander A. Morley***

Department of Haematology and Genetic Pathology, Flinders University and Medical Centre, Bedford Park, South Australia, Australia 5042

Received July 20, 2011; Accepted May 25, 2012



## ABSTRACT

**RNA degradation can distort or prevent measurement of RNA transcripts. A mathematical model for degradation was constructed, based on random RNA damage and exponential polymerase chain reaction (PCR) amplification. Degradation, measured as the number of lesions/base, can be quantified by amplifying several sequences of a reference gene, calculating the regression of $C_t$ on amplicon length and determining the slope. Reverse transcriptase–quantitative PCR (RT–qPCR) data can then be corrected for degradation using lesions/base, amplicon length(s) and the relevant equation obtained from the model. Several predictions of the model were confirmed experimentally; degradation in a sample quantified using the model correlated with degradation quantified using an additional control sample and the ΔΔCt method and application of the model corrected erroneous results for relative quantification resulting from degradation and differences in amplicon length. Compared with RIN, the method was quantitative, simpler, more sensitive and spanned a wider range of RNA damage. The method can use either random or specifically primed complementary DNA and it enables relative and absolute quantification of RNA to be corrected for degradation. The model and method should be applicable to many situations in which RNA is quantified, including quantification of RNA by methods other than nucleic acid amplification.**

## INTRODUCTION

Quantification of messenger RNA (mRNA) by reverse transcription followed by nucleic acid amplification, usually polymerase chain reaction (PCR) [reverse transcriptase–quantitative PCR (RT–qPCR)], is widely performed, but the results are very dependent on the quality of the RNA. This may vary widely, because RNA may be degraded by a variety of physical and chemical factors including heat, radiation, chemicals and tissue ribonucleases. It may be difficult to prevent these, particularly for stored tissue, fixed samples and archival specimens. Several approaches have been used to assess RNA quality. These include spectrophotometry, analysis of 18S and 28S rRNA by electrophoresis, analysis of the complete RNA pattern on electrophoresis (RIN, Agilent Technologies) (1), the 5′–3′ assay (2) and PCR amplification of different target lengths of complementary DNA (cDNA) (3,4). Generally, these methods provide a qualitative or semi-quantitative measure of RNA integrity rather than a quantitative measure, although modification of the RIN algorithm has been claimed to improve RT–qPCR quantification (5), and Gong *et al.* (4) suggested quantitative conclusions from their data. A qualitative measure of integrity may indicate that the sample is adequate for analysis, but a quantitative measure enables quantification of an RNA target to be corrected for degradation.

We recently described a method for quantifying the integrity of genomic DNA in a sample by determining the probability that a base is damaged and the fraction of target molecules that are intact and amplifiable by qPCR. This information enables the true number of target molecules in a sample to be determined (6). The method assumed that DNA damage was random, and the model combined the mathematics of the Poisson distribution and the mathematics of exponential amplification. We have now used a similar approach to model RNA degradation, enabling quantification of RNA integrity. Hence RT–qPCR can be corrected for degradation. Modelling RNA is more complicated owing to the additional step of reverse transcription. Nevertheless, the results are consistent with the model and suggest that the same principles can be used.

*To whom correspondence should be addressed. Tel: +61 4 33101555; Fax: +61 8 82045114; Email: alec.morley@flinders.edu.au

In this article, we describe the mathematical model underlying the method, present experiments testing predictions of the model, illustrate its use for relative quantification and finally discuss practical applications.

## Mathematical analysis

### *Poisson statistics*

A lesion is defined as damage to an RNA or cDNA strand which prevents formation of an intact cDNA target for PCR. The basic assumption of the model is that lesions occur randomly and independently. This assumption seems undoubted when considering external physical or chemical agents, which damage RNA by hydrolytic, phosphorolytic or thermodynamic cleavage, or by the random production of adducts. However, RNA can also be degraded by the action of a large number of ribonucleases, either endoribonucleases or exoribonucleases. Endoribonucleases may show some base or sequence specificity. But, in relation to the total RNA strand, bases and/or short sequences occur at random, so enzyme activity can also be regarded as random. The randomness or non-randomness of exoribonucleases is difficult to assess because there are many enzymes and a variety of mechanisms. However, for most exoribonucleases and for most RNA sequences, the RNA strand degraded by the enzyme is completely degraded (7) and we are not aware of any compelling evidence that this occurs in a non-random fashion. In view of the above considerations, we regard the great majority of RNA degradation as occurring randomly.

In considering RNA degradation in relationship to RT–qPCR, there are three steps during which degradation may influence the relationship between the initial number of RNA targets and the final number of cDNA targets measurable by qPCR. These factors are prior degradation of the RNA in the sample being assayed, degradation of RNA by RT or early termination of cDNA synthesis and degradation of the cDNA. There is an individual probability that a target molecule will survive each of these steps and the overall probability that an initial target molecule will result in a final intact cDNA molecule is the product of the three individual probabilities.

*Degradation of RNA in the sample*. For an RNA target that is to be quantified by RT–qPCR, the target for degradation that needs to be considered is shown in Figure 1. It is the RNA segment that starts at the 3′-RNA base that hybridizes to the most 5′-base of the reverse transcription primer and extends to the 5′-RNA base that corresponds to the 3′-end of the PCR target, in the cDNA. The length of this segment is $(l+p)$ bases, where $l$ is the length of the PCR target and $p$ is the length of the cDNA strand from its 5′-end to the beginning of the PCR target. For random priming, $p$ is a mean value.

The probability that a given number of lesions will affect the RNA segment is described by the binomial distribution. If the mean number of lesions/base in RNA is $r_1$, then the probability that there will be no lesions affecting the RNA segment is $(1-r_1)^{(l+p)}$. When $r_1$ is very small, the Poisson distribution provides a good approximation to the binomial distribution. The probability $P_1(0)$ of no lesions in the segment is the zero term of a Poisson distribution where $\mu$ is the mean number of lesions in the strand. Thus,

$$P_1(0) = \mathrm{e}^{-\mu} = \mathrm{e}^{-(l+p)r1}.$$

*Degradation of RNA by RT*. RT has both polymerase and ribonuclease activity. A stem-loop secondary structure in the RNA template may block polymerase extension, but this only occurs with occasional targets and is minimized by the use of a RT, such as Superscript III, which can operate at a relatively high temperature. Impaired conversion of RNA targets to cDNA targets as a result of general impairment of processivity is unlikely, given the processivity of the enzyme (2400 bases/min for Superscripts II and III) and the long duration of the RT phase of RT–qPCR (30–40 min). Degradation of RNA by the RNase activity of RT is a possibility. Naturally occurring RTs all have RNase H activity, but this has been decreased or eliminated in genetically engineered forms of the enzyme which are now widely used when performing RT–qPCR. Although degradation of RNA by RT is unlikely to be of major importance, we have incorporated it into our model by introducing the term $r_2$, the probability/base that RT will cause RNA strand breakage and thus failure to complete cDNA synthesis. Again, as in the previous section (a), the probability $P(0)$ that an intact RNA strand will be converted into an intact cDNA target during reverse transcription is given by

$$P_2(0) = e^{-(l+p)r2}.$$

*Degradation of cDNA*. If $r_3$ is the probability/base that the cDNA strand will suffer a lesion, which prevents polymerase amplification, then, as in the previous sections, the probability $P(0)$ that a synthesized cDNA strand will persist as an intact target for PCR is given by:

$$P_3(0) = e^{-lr3}.$$

However, degradation of cDNA prior to performance of qPCR is unlikely to be quantitatively important, owing to the relative stability of DNA and as qPCR is usually performed immediately after the reverse transcription

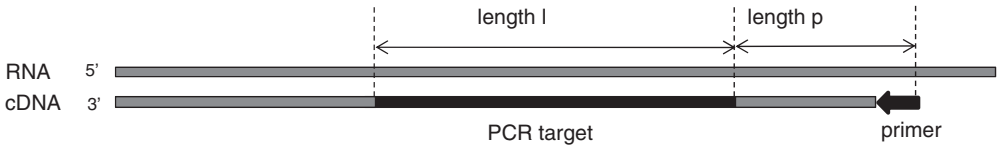


**Figure 1.** The regions of RNA and cDNA used for the mathematical model, and the locations of variables $l$ and $p$, used in that model. $l$ is the length of the PCR product; $p$ is the length of the region of cDNA, which extends from its 5′-end to the upstream PCR primer binding site.

step. For this reason $r_3$ approximates 0 and $P_3(0)$ approximates 1.

Multiplying $P_1(0)$, $P_2(0)$ and $P_3(0)$ gives the overall probability $P(0)$ that an original RNA target will be converted to a cDNA target that is amplifiable and quantifiable by the qPCR. Thus,

$$P(0) = e^{-(l+p)\cdot(r1+r2)} = e^{-(l+p)r},$$

where $r = r_1 + r_2$

If the amplifiable fraction (AF) is the proportion of RNA target molecules that are undamaged and which result in corresponding cDNA molecules amplifiable in the PCR, then

$$\mathrm{AF} = P(0) = e^{-(l+p)r} \quad (1)$$

Thus, for any RNA sample, $AF$ of a target can be calculated if $l$ and $p$ are known and by measuring $r$. The method for measuring $r$ is described below.

***Exponential amplification***

During exponential amplification

$$N_c = N_o a^c,$$

where $a$ is the amplification efficiency, defined as the proportional increase per cycle; $c$ is the number of cycles; $N_0$ is the initial number of sequences being amplified and $N_c$ is the number of amplified sequences present after $c$ cycles. When $c = C_t$, the threshold number of cycles for real-time PCR

$$N_{Ct} = N_0 a^{Ct}. \quad (2)$$

If $m$ is the mass of cDNA initiating the PCR and $z$ is the mass containing one instance of the target sequence,

$$N_{Ct} = (m/z) \cdot AF \cdot a^{Ct}. \quad (3)$$

From Equations (1) and (3)

$$N_{Ct} = m/z.e^{-(l+p)r} \cdot a^{Ct}.$$

Rearranging and taking logarithms to base $e$

$$C_t \cdot \log_e a = (l+p)r + \log_e(N_{Ct} \cdot z/m)$$

Since $N_{Ct}$, $z$ and $m$ are constants, the equation has the following form

$$C_t \cdot \log_e a = (l+p)r + \text{constant}. \quad (4)$$

There is therefore a linear relationship between $C_t \log_e a$ and $(l+p)$, with slope $r$. When random priming is used for reverse transcription, $p$ has a mean value that is the reciprocal of the probability that priming will commence at any one base. The value of $r$ can then be directly calculated as the slope of the linear relationship between $C_t \log_e a$ and $l$. When gene-specific or poly-dT priming is used, the value of $p$ for each length is known, and the value of $r$ can then be calculated as the slope of the linear relationship between $C_t \log_e a$ and $(l+p)$. When random priming is used, the value of $p$ can be determined by using Equation (4), which indicates that, when $l$ equals $-p$, the value of $C_t$ becomes constant and independent of $r$. This enables the value of $p$ to be determined by studying RNA samples degraded to different extents (see 'Results' section and Figure 2).

Although the value of $a$ can be empirically determined for each sequence length, it is much simpler to use a robust, efficient PCR amplification system, for which the value of $a$ is known and is constant over the range of sequence lengths studied. Optimization of the PCR conditions enables this be achieved for a reasonable range of amplicon sizes. If $a$ is constant, then the slope of $C_t$ versus $l$, or $C_t$ versus $(l+p)$, can be determined, and

$$r = slope \cdot log_e a. \quad (5)$$

In most situations, RNA targets of interest are quantified relative to a standard gene in the same sample. If, in the sample, $N_{tar}$ is the initial total number of molecules, intact and degraded, of a target to be quantified, then from Equations (1) and (2)

$$N_{Ct} = N_{tar} \cdot e^{-(l+p)r}.a^{Ct}. \quad (6)$$

The assumptions for relative quantification of a target sequence (subscript tar) against a standard reference or 'housekeeping' sequence (subscript st) are that $N_{Ct}$ and $a$ are the same for both target and standard. Since both target and standard are quantified in the same RNA/cDNA sample, $r$ is also the same. If $p$ is also the same, as is the case for random priming, then

$$N_{tar}/N_{st} = e^{-(l_{st}-l_{tar})\cdot r} \cdot a^{(Ct_{st}-Ct_{tar})}. \quad (7)$$

This is the familiar $\Delta\Delta C_t$ calculation (8), corrected for degradation which takes account of product length. For gene-specific or oligo dT priming, $l$ needs to be replaced by $l+p$. The value of $r$ is most simply obtained by amplifying several different lengths of a reference sequence in the test sample and applying Equation (5). The reference sequences can be the same gene as the target, but is most conveniently a separate sequence for which amplification efficiency has been previously well characterized. If relative quantification is performed using the geometric mean of results for several standards (9), then $l_{st}$ is the arithmetic mean of the lengths of those standards.

In the uncommon situation where an RNA sequence in a test sample (subscript t) is quantified relative to the same RNA sequence in an external reference sample (subscript e), $r_t$ and $r_e$ must be determined. This can be done by amplifying several lengths of the sequence of interest, or of a separate reference sequence, in each sample and applying Equation (5). The ratio of AF of the target sequence in the test sample, to AF in the external samples, can then be determined, because from Equation (1)

$$\mathrm{AF_t/AF_e} = e^{-(l+p).(r_t-r_e)}.$$

This ratio can then be used to correct the RT–qPCR results for degradation.

With gene-specific or oligo-dT priming, the value of $r$ can also be obtained by amplifying two sequences, one at the 3′-end and one at the 5′-end of a cDNA sequence. Assuming that the number of amplicons at threshold is

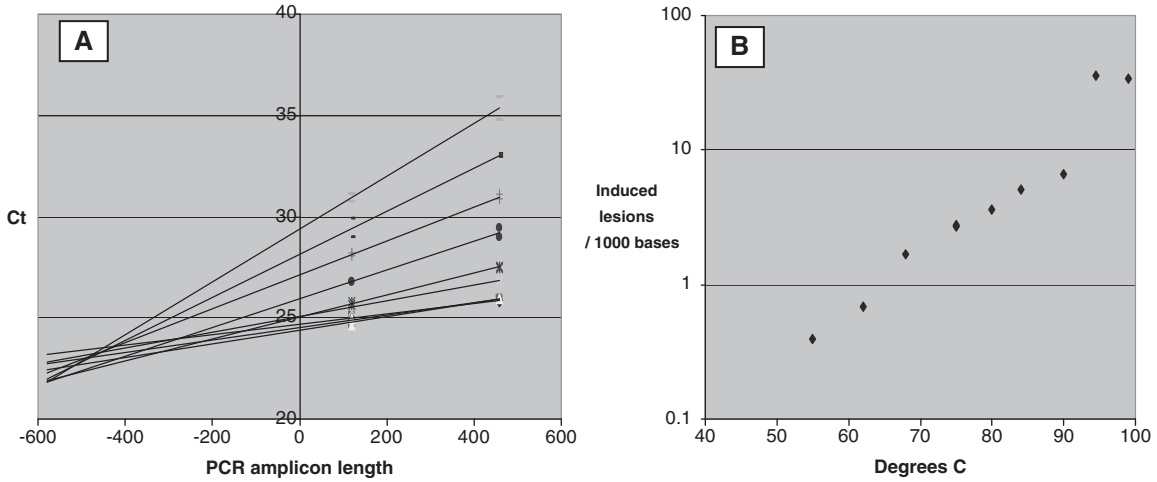


**Figure 2.** RNA degradation by heat. Nine RNA samples were held at different temperatures for 30 min and then analysed by RT-qPCR. (**A**) The $C_t$ values for the different samples. (**B**) The calculated values for lesions per 1000 bases. In A, the regression lines for the $C_t$ values have been extrapolated leftwards to estimate the value of $p$. This value equals $-l$, where $l$ is the amplicon length at which the regression lines intersect. In this experiment, which used random priming for reverse transcription, intersects mostly fell between −300 and −500 bases.

the same for the two probes, rearrangement of Equation (6) leads to the relation

$$r = (log_e a \cdot (Ct_{3'} - Ct_{5'}))/((l_{3'}+p_{3'}) - (l_{5'}+p_{5'})),$$

where the subscripts 3′ and 5′ refer to the PCR amplifications of the sequences at the 3′- and 5′-regions of the cDNA. This equation reduces to

$$r = (log_e a \cdot (Ct_{3'} - Ct_{5'}))/(l_{3'}+l_i), \quad (8)$$

where $l_i$ is the length of the sequence intervening between the most 5′-base of the 3′-sequence and the most 3′-base of the 5′-sequence.

## MATERIALS AND METHODS

Blood samples were collected from a healthy volunteer, into EDTA, and RNA was extracted immediately. RNA extraction used the QiaAmp RNA Mini Kit (Qiagen). In brief, erythrocytes were lysed in hypotonic media; leukocytes were pelleted, then lysed and nucleic acids precipitated with ethanol; RNA was bound to a spin column; the column was washed, then RNA was eluted in sterile RNAse-free water. RNA was either used immediately or aliquotted and frozen at −80°C. Aliquots, once thawed, were not re-frozen.

Unless stated, all studies used the *NRAS* gene (neuroblastoma RAS viral oncogene homologue, geneID 4893, MIM 164790). Total RNA was reverse transcribed with Superscript III (Invitrogen) and RNAseOUT recombinant ribonuclease inhibitor (Invitrogen), with either specific primers or random hexamers, according to the manufacturer's protocol. Primers were first annealed to RNA by heating to 65°C for 5 min and then cooling on ice for at least 1 min. Reverse transcription was in 25 $\mu$l of 50 mM Tris–HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT, 40 units RNAseOUT, 200 units Superscript III, 0.5 mM each dNTP and total RNA from up to 120 $\mu$l blood (i.e. 500,000 leukocytes). If random primers were used, the mixture was then incubated at 25°C for 5 min. Reverse transcription was at 50°C for 30 min. After which, the reaction was stopped by heating to 70°C for 15 min. cDNA was used immediately or stored at −80°C.

PCRs were performed in 25 $\mu$l volumes containing 1–2 U Platinum Taq (Invitrogen), 20 mM Tris–HCl pH 8.4, 50 mM KCl, 5 mM $MgCl_2$, 300 $\mu$M each of dATP, dGTP, dCTP or dUTP; 50 ng each primer; up to 0.4 $\mu$l cDNA and 20 ng (2–2.5 pmol) hydrolysis probe. Reactions were manually set up in 96-well 0.2 ml Axygen PCR microplates with strip clear flat top caps (Axygen, Union City, CA, USA). Cycling conditions were as follows: an initial denaturation at 96°C for 2 min; then up to 55 cycles of 94°C for 15 s; 55°C for 90 s and 72°C for 60 s. Cycling was done on a BioRad IQ5 thermal cycler, running IQ5 V2.0.148.60623 software. The features felt to be of particular importance for efficient amplification were the high concentrations of $MgCl_2$ and dNTPs and the prolonged annealing time.

The primers used are shown in Table 1. To generate a set of PCR products of varying lengths—all RT–qPCRs used the same probe, and the same primer 5′ of the probe (extension of the primer hydrolyses the probe). The remaining primers were designed to give amplicons between 80 and 400 bases in length. Primers were aimed for a $T_m$ of >51°C. (nearest neighbor method) under PCR reaction conditions.For the RIN work, duplicate aliquots of RNA were incubated at temperatures between 55 and 92°C for 30 min, and then one aliquot was reverse transcribed and analysed as above; the other analysed by an Agilent 2100 Bioanalyser (Agilent Technologies), with an RNA 6000 Pico Labchip Kit. Amplicon sizes: sizes stated here include the primers.

## RESULTS

In the previous study of DNA degradation, using an optimized amplification protocol, the value of $a$ was

**Table 1.** primer sets for assessing degradation in RNA/cDNA

| | | |
|---|---|---|
| *NRAS* | probe | VIC-attgtcagtgcgctttttcccaacaccacctg-BHQ |
| | F1 | cacttgttttctgtaagaatcctct |
| | F2 | actcgcttaatctgctccctgt |
| | F3 | aaaaagcatcttcaacaccctgtc |
| | F4 | tcttttctgacaaaactttaaaagtatcttg |
| | F5 | gaaatgactgagtacaaactggtggtg |
| | F6 | ccggctgtggtcctaaatctgtc |
| | F7 | gggctgttcatggcggttcc |
| | R1 | aggcagtggagcttgaggttc |
| | R2 | gaatcctctatggtgggatcatattcatct |
| *GAPDH* | Probe 1 | FAM-ccatcaccatcttccaggagcgagatccctc-BHQ |
| | Probe 2 | FAM-ccatccacagtcttctgggtggcagtgatg-BHQ |
| | F1 | gagaacgggaagcttgtcatcaatgg |
| | F2 | gacccctcattgacctcaactacatgg |
| | F3 | tgcaggggggagccaaaagg |
| | F4 | cagcctcaagatcatcagca |
| | F5 | ccatgacaactttggtatcg |
| | F6 | gtggaaggactcatgaccac |
| | R1 | ccagcatcgccccacttgatt |
| | R2 | atggtggtgaagacgccagtgg |
| | R3 | ggggcagagatgatgacccttttgg |
| | R4 | ggttcacacccatgacgaacatgg |
| | R5 | aggaggcattgctgatgatcttgagg |
| | R6 | gtccttccacgataccaaagttgtcatgg |
| | R7 | cacgccacagtttcccggag |
| *APC* | Probe 1 | FAM-aagcagaattagatgctcagcactt-BHQ |
| | probe 2 | FAM-aagtgctgagcatctaattctgctt-BHQ |
| | F1 | tgccatctcttcatgttagg |
| | F2 | ggaatctcatggcaaatagg |
| | F3 | gccaatattatgtctcctgg |
| | F4 | gcacaaaatgattgctatgg |
| | F5 | ggaagcattatgggacatgg |
| | R1 | actacgatgagatgccttgg |
| | R2 | catcatgtcgattggtgtc |
| | R3 | aaggacagtcatgttgccag |
| | R4 | ttcctcttgatgaagaggag |
| | R5 | ctagaccaattccgcgttct |
| | R6 | gccttgggacttaaattgtc |

found to be constant over the range of lengths studied for the *NRAS* sequence and for four other sequences (6). This study also used the same amplification protocol and the *NRAS* gene and the same probe, but different target sequences, and primers on exon–exon boundaries.

In four studies involving products with lengths from 120 to 458 bp, the mean amplification efficiency/cycle was 1.89 and the regression line between amplification efficiency and length had a slope of $-5.2 \times 10^{-5}$, which was not significantly different from zero.

The number of lesions per 1000 bases was determined in 33 studies, 9 on freshly prepared RNA and 24 on RNA that had been frozen at −80°C and thawed for study. In the initial 12 studies, the regression line of $C_t$ versus amplicon length was determined using at least three different amplicons, but subsequent studies generally used only two amplicons. The mean value for lesions per 1000 bases was 1.47 for the fresh samples and 1.63 for the frozen samples but the difference was not significant ($t$-test, $P > 0.05$) and the results were pooled. The mean (SD) value for the 33 studies was 1.59 (0.61) and the mean value was significantly different from 0 ($t = 2.58$, $P < 0.01$, one-tailed $t$-test).

In a total of 22 studies, RNA degradation was produced by heating RNA for 30 min, and in every case an increase in the measured value of degradation was observed. In each of the three experiments, nine different temperatures were used, and in each case increasing temperature produced increasing degradation; the results from one of the three experiments are shown in Figure 2. The left panel shows the various regression lines for the relationship between $C_t$ and $l$, and the right panel shows the calculated numbers of lesions/base, in relation to temperature. In these studies, reverse transcription was initiated by random priming and thus the value of $p$ was not known *a priori*. However, our model indicates that, when $l$ equals $-p$, the $C_t$ value should be independent of $r$, and it therefore predicts that the various regression lines should intersect at around the value of $-p$. This prediction was observed in each study in which RNA was degraded by heating. The points of intersection were between −200 and −500 bases, although there was marked experimental variation, presumably owing to variation in $C_t$ results having a marked effect on the slopes of the regression lines.

To further test the model, we experimentally manipulated $p$ and observed the effect on the point of intersection between the regression line for control RNA, and the regression line for RNA degraded by heating. This was done in two ways. (i) Manipulation of random priming was performed in four experiments. Since the probability of transcription being initiated at any one base is the reciprocal of $p$, we reasoned that varying the concentration of random primers would influence the probability of the initiation of reverse transcription and hence the value of $p$ as determined from the intersection point would decrease as the primer concentration increased. Of the four experiments, one used 8, 60 and 200 ng of hexamers; the other three used 4, 20 and 100 ng hexamers. Regression lines were based on the results from amplicons of two lengths, each amplified in triplicate. For the four experiments, the estimated values of $p$ were: 224, 107 and 97 bp; 286, 283 and 161 bp; 1741, 187 and −57 bp; 422 214 and 174 bp. Despite the experimental variation, in each experiment, an increasing concentration of hexamers produced a decreasing estimate for $p$. (ii) Manipulation of gene-specific priming was performed in one experiment. Three values of $p$ were produced by using three reverse transcription primers, and for each primer, a series of PCR primers was used in order to determine the intersection point between the regression lines for control and degraded DNA and thus estimate the value of $p$ for that transcription primer. Three amplicons were PCR amplified each in triplicate. The expected values for $p$ were 0, 220 and 338 bp; the observed results were 8, 118 and 408 bp. Thus the results of the two approaches to manipulating $p$ were consistent with the predictions of the model.

In another four experiments using control and heated RNA (data not shown), various combinations of gene-specific primers were used for initiation of reverse transcription to produce various values for $(l+p)$. These experiments directly examined Equation (3), which predicts a linear relationship between $C_t$ and $(l+p)$ and predicts an intersection point when $(l+p)$ equals zero. The four intersection points were at−14, −13, 32 and 119 bases. Given the imprecision of measurement of the

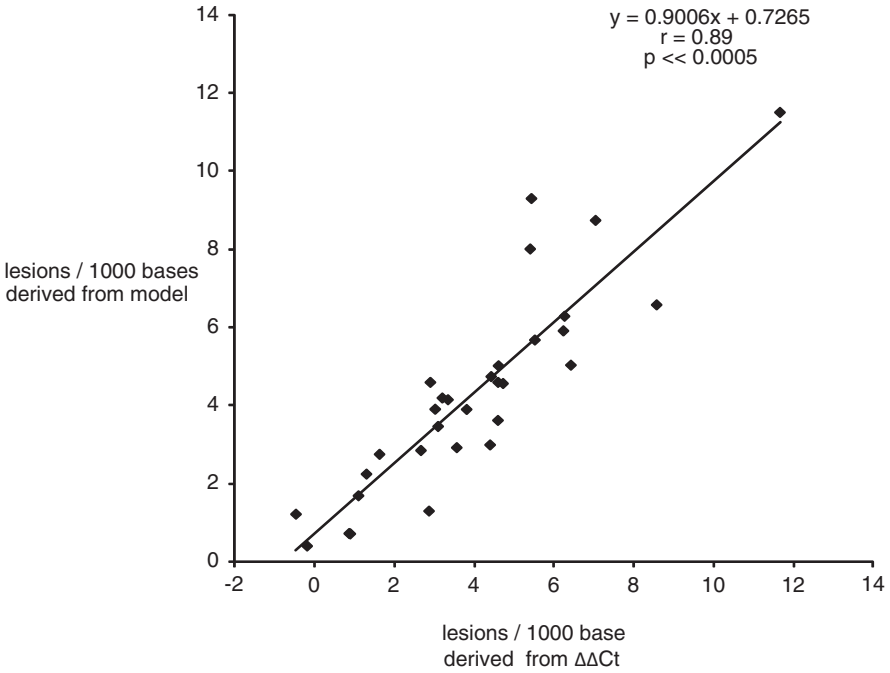


**Figure 3.** Correlation between measuring RNA integrity by the present method and by the $\Delta\Delta C_t$ method. The RNA in 30 samples was degraded by heat. The present method quantified degradation by studying the degraded sample only, whereas the $\Delta\Delta C_t$ method quantified degradation using the $C_t$ difference between the degraded sample and a control undegraded sample of the same RNA.

intersection point, these results are consistent with the prediction of the model.

When a control sample of RNA is available, the extra lesions produced by heating can be determined by RT–qPCR, by amplifying the same sequence in both samples, noting the $C_t$ difference between the control and heated sample and calculating using the $\Delta\Delta C_t$ method. From the results of the various experiments in which RNA was degraded by heat and in which the control sample was undegraded RNA, it was possible to compare for each heated sample, the lesions per base as determined by the $\Delta\Delta C_t$ method with the lesions per base as determined by using Equation (5) of our model. The results of this comparison, shown in Figure 3, indicate a highly significant correlation (lesions per base: model $= 0.9006 \times [\Delta\Delta C_t] + 0.73$; $r = 0.89$, $P < 0.0005$).

In one experiment, the present method was compared with the RIN method for quantification of RNA damage produced by heating. The results are shown in Figure 4. Except possibly for a small window, the RIN method suggested that the RNA was either completely intact or completely damaged, whereas the present method showed progressively increasing RNA degradation as the degree of heating was increased. In the samples that RIN indicated were completely damaged, the present method was able to distinguish different degrees of damage and indicated that most small targets were still intact.

To determine whether our model could be applied to other gene targets, we investigated two other genes, GAPDH (glyceraldehyde-3-phosphate dehydrogenase; Gene ID2597; MIM138400) and APC (adenomatous polyposis coli, geneID 324, MIM611731) and for each gene, synthesized two series of primers, with each series

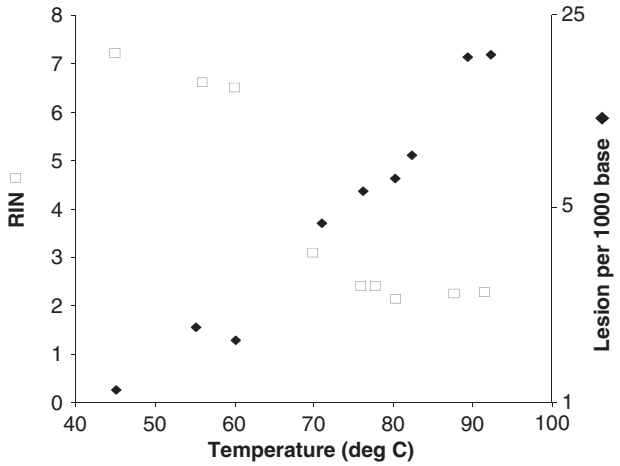


**Figure 4.** Comparison between the method described here for measuring lesions/base, and the RIN score, which is based on automated electrophoresis and which gives a result between 1 (totally degraded) and 10 (fully intact). Nine RNA samples were partially degraded by heating at various temperatures for 30 min and then analysed. The number of lesions/base shows a progressive increase as the temperature of heating increases whereas RIN shows a largely 'all-or-nothing' response.

comprising 1 upstream primer and 5 or 6 downstream primers. Amplicon sizes range overall, from 63 to 417 bp. Within each series, the smallest amplicon was between 63 and 182 bases; and largest amplicon was 155–249 bases longer. When the integrity of RNA in a sample was investigated with these four series of primers: the mean amplification per cycle was 1.97, 1.91,

**Table 2.** Relative quantification of different lengths of *GAPDH* mRNA relative to different lengths of *APC* mRNA as the standard

| | | conventional relative quantification | | | | corrected relative quantification | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | GAPDH (bp) | | | | GAPDH (bp) | | | |
| | APC (bp) | 81 | 100 | 166 | 257 | 81 | 100 | 166 | 257 |
| expt 1 | 63 | 4.9 | 4.6 | 5.6 | 4.6 | 4.9 | 4.7 | 6.1 | 5.3 |
| control RNA | 85 | 8.8 | 8.3 | 10.1 | 8.3 | 8.7 | 8.4 | 10.7 | 9.4 |
| 0.7 lesions / 1000 bases | 161 | 9.2 | 8.7 | 10.6 | 8.7 | 8.6 | 8.3 | 10.7 | 9.3 |
| | 218 | 7.1 | 6.7 | 8.2 | 6.7 | 6.4 | 6.2 | 7.9 | 6.9 |
| degraded RNA | 63 | 11.2 | 7.5 | 6.0 | 3.5 | 12.6 | 9.4 | 11.4 | 11.9 |
| 6.3 lesions / 1000 bases | 85 | 14.7 | 9.8 | 7.8 | 4.6 | 14.4 | 10.8 | 13.0 | 13.6 |
| | 161 | 34.8 | 23.2 | 18.5 | 10.9 | 21.0 | 15.8 | 19.1 | 19.9 |
| | 218 | 28.9 | 19.2 | 15.3 | 9.0 | 12.2 | 9.1 | 11.0 | 11.5 |
| expt 2 | 63 | 3.9 | 3.8 | 2.6 | 2.8 | 4.1 | 4.2 | 3.3 | 4.5 |
| control RNA | 85 | 7.8 | 7.6 | 5.1 | 5.7 | 7.7 | 7.8 | 6.3 | 8.6 |
| 2.4 lesions / 1000 bases | 161 | 9.2 | 9.0 | 6.1 | 6.7 | 7.6 | 7.8 | 6.2 | 8.5 |
| | 218 | 6.8 | 6.7 | 4.5 | 5.0 | 4.9 | 5.0 | 4.0 | 5.5 |
| degraded RNA | 63 | 5.7 | 4.3 | 2.3 | 1.1 | 6.5 | 5.7 | 5.0 | 4.7 |
| 7.6 lesions / 1000 bases | 85 | 10.8 | 8.2 | 4.4 | 2.0 | 10.4 | 9.1 | 8.1 | 7.5 |
| | 161 | 20.3 | 15.4 | 8.3 | 3.9 | 11.1 | 9.7 | 8.6 | 8.0 |
| | 218 | 14.7 | 11.2 | 6.0 | 2.8 | 5.2 | 4.6 | 4.0 | 3.8 |
| expt 3 | 63 | 8.4 | 6.8 | 7.8 | 6.8 | 8.5 | 7.2 | 8.9 | 8.5 |
| control RNA | 85 | 13.4 | 11.0 | 12.6 | 10.9 | 13.3 | 11.2 | 13.9 | 13.4 |
| 1.2 lesions / 1000 bases | 161 | 19.7 | 16.1 | 18.5 | 16.0 | 17.9 | 15.0 | 18.6 | 17.9 |
| | 218 | 12.0 | 9.8 | 11.2 | 9.7 | 10.2 | 8.5 | 10.6 | 10.2 |
| degraded RNA | 63 | 15.2 | 10.1 | 8.2 | 3.7 | 17.2 | 13.1 | 16.9 | 14.2 |
| 7.0 lesions / 1000 bases | 85 | 24.6 | 16.3 | 13.3 | 5.9 | 23.9 | 18.1 | 23.5 | 19.7 |
| | 161 | 50.8 | 33.7 | 27.5 | 12.2 | 29.0 | 22.0 | 28.4 | 23.9 |
| | 218 | 42.3 | 28.0 | 22.9 | 10.2 | 16.2 | 12.3 | 15.9 | 13.4 |

Relative quantification was performed either by the conventional method or after quantifying degradation and then applying Equation (7). The RNA was either control RNA or RNA degraded by heating at 91°C for 30 min. The results obtained by conventional relative quantification are influenced by the length of the test amplicon, the length of the standard amplicon and the presence of degradation. These effects disappear after correction for degradation and length. For degraded RNA, in each experiment correcting the results of conventional relative quantification resulted in a highly significant ($P < 0.0005$) decrease in their variance; for control RNA correction produced a decrease in variance in each case but the decreases were not significant.

2.07 and 2.07, respectively; the slopes of the relationship between amplification efficiency and amplicon length were $-6.1 \times 10^{-5}$, $3.3 \times 10^{-5}$, $-4.0 \times 10^{-4}$ and $-8.2 \times 10^{-4}$, respectively, with none of these values being significantly different from 0; and the calculated lesions per 1000 bases were 5.6, 4.6, 5.5 and 3.1, respectively.

In six further experiments, we used the three genes to measure lesions per 1000 bases in control RNA and RNA degraded by heating and to calculate relative quantification of each amplicon of two genes using each amplicon of the third gene in turn as a standard. Relative quantification was calculated either conventionally or by using our model. For application of our model, degradation was measured by applying Equation (5) to the $C_t$ values obtained from amplification of the various amplicons of each gene, and relative quantification was then calculated using Equation (7). The results in all experiments were similar and were not affected by which gene was used for test RNA and which for the standard. For conventional relative quantification, if the standard amplicon was constant then the result for the test amplicon decreased with increasing amplicon length and, conversely, if the test amplicon was constant then the result increased as the length of the standard amplicon increased. These effects were increased by degradation but were absent when our model was used to calculate relative quantification. The results of quantification of GAPDH relative to APC from three consecutive experiments are shown in Table 2.

To determine the minimum amount of RNA needed for our method, we performed experiments each using serial dilution of RNA, reverse transcription using random or gene-specific priming and assessment of degradation. GAPDH was used in these experiments because it had the highest RNA level. The results of these experiments are shown in Table 3. For undegraded RNA quantification of integrity required ~1 pg of RNA; for degraded RNA, the amount of RNA required increased as the extent of degradation increased. For degraded RNA, less RNA was required when gene-specific priming was used, presumably because the zero value of $p$ resulted in less degradation.

## DISCUSSION

The model provides a simple method for quantifying RNA integrity and hence improving quantification of RNA by RT–qPCR. Quantification of RNA integrity is

**Table 3.** The least mass of RNA required for quantification of degradation

| Degradation degree Celsius for 30 min | Priming | Minimum pg RNA for quantification | Lesions/1000 bases | |
|---|---|---|---|---|
| | | | Best estimate | Estimate at limit |
| Control | Random | <16 | 1.1 | −0.2 |
| Control | Random | 1 | 1.4 | 1.4 |
| 85 | Random | 16 | 3.9 | 2.3 |
| 88 | Random | 63 | 7.5 | 7.8 |
| 91 | Random | 126 | 9.7 | 27.9 |
| Control | Gene-specific | 5.3 | 2.8 | 4.5 |
| 91 | Gene-specific | 10 | 9.7 | 9.0 |
| 91 | Gene-specific | 10 | 9.4 | 4.3 |

based on amplifying different lengths of cDNA sequences of a reference transcript. The method assumes that the amplification efficiency is independent of length and is known. Our protocol for fulfilling this assumption for genomic DNA has been reported (6) and is also used here, but any protocol which fulfils this assumption would be suitable.

Although the model covers quantification of RNA using an external sample, in nearly all cases target RNA is quantified relative to one or more standard RNAs in the same sample. The model enables RNA integrity to be determined by PCR amplification of several lengths of a reference sequence, using the same probe and with application of Equation (5). Little extra work is involved particularly for an experiment in which multiple RNA species are being quantified. Once RNA integrity has been quantified, target RNA sequences can be quantified, by applying Equation (7). If desired, an RNA standard can also be used as a reference providing that the amplification characteristics of different lengths of the standard are known. A panel of RNA reference genes, rather than a single reference gene, can also be used.

Equation (7) indicates that correction for RNA degradation becomes increasingly important as the difference between the length of the target and the reference increases, and/or as the degree of degradation increases. (Here, length is $l$ for random priming and $l+p$ for gene-specific or oligo dT priming). This theoretical prediction was confirmed experimentally as the effects of length and degradation were observed when relative quantification was calculated conventionally but were not observed when relative quantification was calculated according to our model (see 'Results' section and Table 2). The experimental results and the underlying theory indicate the desirability of locating the primers for the target and standard, so that $l$ or $l+p$ are similar for each. At times, in addition to quantification, the absolute level of detection of an RNA may also be important. The model suggests that minimizing $l+p$ will improve detection. The value of $l$ can be decreased by using short amplicons. The value of $p$ can be decreased by improving hybridization of the reverse transcription primer, either by increasing its concentration or increasing its $T_m$, or by using a gene-specific primer.

Current methods for assessing RNA degradation only give a qualitative or semiquantitative result and are generally used only to decide whether the sample is sufficiently intact to enable further analysis. Perhaps the most widely used current approach for assessing RNA degradation is RIN, which assesses mRNA indirectly, by assessing ribosomal RNA. Our method has a number of advantages over RIN since: it gives a quantitative result which enables RT–qPCR results to be corrected for degradation; it enables gene expression to be quantified in samples which RIN would suggest were too degraded to be analysable; it requires over an order of magnitude less RNA than RIN; and it is simpler, cheaper and does not require a dedicated instrument.

A recent detailed review (10) of the effects of RNA degradation on the results of gene expression studies, concluded that RNA quality influences results and that some improvement could be obtained by performing relative quantification using the geometric mean of 4 standard housekeeping genes or by performing a 5′–3′ assay on cDNA produced by oligo dT priming. Our model is compatible with either one or a number of genes as standards but, as indicated in Equation (7), length and degradation both need to be considered. The 5′–3′ assay as originally described (2) provides only a semiquantitative indication of degradation, but Equation (8) indicates that it can provide a truly quantitative measure of degradation, which can then be used in Equation (7) to calculate relative quantification. The 5′–3′ assay has the advantage that the amplicons for qPCR can be quite short, which minimizes the risk of decreased amplification efficiency associated with long amplification, but the method assumes that the fluorescence thresholds for the two probes will be reached at the same number of amplified molecules. Furthermore, when the experiment involves random priming, it would seem simpler to determine $r$ by amplifying several lengths of a reference sequence, and thus avoid having to perform additional oligo dT or gene-specific priming.

Our model is based on several assumptions, which are not unique to it. RNA degradation may not always be random; the efficiency of random priming may not be the same for all genes; the efficiency of amplification and the number of amplicons at threshold may not be the same for all genes quantified; and secondary structure may affect reverse transcription and cause under-estimation of transcript number. Nevertheless, these assumptions seem reasonable approximations and, provided they are borne in mind, we believe that our model will provide a useful tool for quantifying RNA integrity and improving the quantification of gene expression.

Although our model has been constructed and tested from the perspective of PCR, mRNA can be quantified by a number of methods which do not involve nucleic acid amplification. For any method, there will be a critical region of RNA, of length $L$ bases, which is characterized by the property that a lesion of any base within it will affect quantification whereas a lesion of any base outside it will have no effect. For PCR, $L$ is the same as $l+p$ . Our model will apply to any method in which the value of $L$ varies for different genes. Several methods involve

hybridization at one point along the RNA strand and quantification of another region at a variable distance along the RNA strand. For example: for microarray, length variation will occur if the labelled cDNA is produced by oligo-dT priming or if there is any variation in the length of the hybridization probes; for SAGE and related methods, length variation will occur owing to variation between the point of oligo-dT capture and the point of restriction enzyme digestion; for Nanostring, length variation will occur if there is variation between the points of hybridization of the capture probe and the measurement probe; and for RNA-Seq, if capture by oligo-dT is used, prior degradation will influence the frequency of reads along the RNA strand. With each method, our PCR method may be used to quantify RNA degradation and the results of relative quantification may then be corrected using Equation (7) and substituting $L$ for $l$ in that equation. RNA-Seq also enables RNA degradation to be measured directly, by measuring the frequency with which reads are recovered along the cDNA strand or, alternatively, degradation does not need to be separately measured if quantification for all genes is performed by determining read frequency at the same distance from the polyA region.

## FUNDING

Monoquant Proprietary Limited. Funding for open access charge: Monoquant Proprietary Limited to Flinders University.

*Conflict of interest statement*: A.A.M. and M.J.B. hold equity in Monoquant which has lodged a provisional patent covering the method.

## REFERENCES

1. Fleige,S. and Pfaffl,M.W. (2006) RNA integrity and the effect on the real-time qRT-PCR performance. *Mol. Aspects Med.*, **27**, 126–139.
2. Nolan,T., Hands,R.E. and Bustin,S.A. (2006) Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.*, **1**, 1559–1582.
3. Bauer,M., Polzin,S. and Patzelt,D. (2003) Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of bloodstains? *Forensic Sci. Int.*, **138**, 94–103.
4. Gong,X., Tao,R. and Li,Z. (2006) Quantification of RNA damage by reverse transcription polymerase chain reactions. *Anal. Biochem.*, **357**, 58–67.
5. Ho-Pun-Cheung,A., Bascoul-Mollevi,C., Assenat,E., Boissiere-Michot,F., Bibeau,F., Cellier,D., Ychou,M. and Lopez-Crapez,E. (2009) Reverse transcription-quantitative polymerase chain reaction: description of a RIN-based algorithm for accurate data normalization. *BMC Mol. Biol.*, **10**, 31–40.
6. Brisco,M.J., Latham,S., Bartley,P.A. and Morley,A.A. (2010) Incorporation of measurement of DNA integrity into qPCR assays. *Biotechniques*, **49**, 893–897.
7. Houseley,J. and Tollervey,D. (2009) The many pathways of RNA degradation. *Cell*, **136**, 763–776.
8. Schmittgen,T.D. and Livak,K.J. (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.*, **3**, 1101–1108.
9. Vandesompele,J., De Preter,K., Pattyn,F., Poppe,B., Van Roy,N., De Paepe,A. and Speleman,F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**, RESEARCH0034.
10. Vermeulen,J., De Preter,K., Lefever,S., Nuytens,J., De Vloed,F., Derveaux,S., Hellemans,J., Speleman,F. and Vandesompele,J. (2011) Measurable impact of RNA quality on gene expression results from quantitative PCR. *Nucleic Acids Res.*, **39**, e63.