Biomolecular Detection and Quantification 12 (2017) 1–6

Contents lists available at ScienceDirect

Biomolecular Detection and Quantification

journal homepage: www.elsevier.com/locate/bdq

Research paper

Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR)

Amin Forootan[a,b,*], Robert Sjöback[c], Jens Björkman[c], Björn Sjögreen[b], Lucas Linz[d], Mikael Kubista[c,e]

[a] Sahlgrenska Cancer Center, Department of Pathology, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
[b] MultiD Analysis AB, Gothenburg, Sweden
[c] TATAA Biocenter AB, Gothenburg, Sweden
[d] LGC Douglas Scientific, USA
[e] Institute of Biotechnology, CAS, BIOCEV, Vestec, Czech Republic

ARTICLE INFO

Handled by Jim Huggett

Keywords:
Limit of detection
Limit of quantification
LoD
LoQ
GenEx software
Real-time PCR
qPCR
Data analysis
Replicates
MIQE
Quality control
Standardization

ABSTRACT

Quantitative Real-Time Polymerase Chain Reaction, better known as qPCR, is the most sensitive and specific technique we have for the detection of nucleic acids. Even though it has been around for more than 30 years and is preferred in research applications, it has yet to win broad acceptance in routine practice. This requires a means to unambiguously assess the performance of specific qPCR analyses. Here we present methods to determine the limit of detection (LoD) and the limit of quantification (LoQ) as applicable to qPCR. These are based on standard statistical methods as recommended by regulatory bodies adapted to qPCR and complemented with a novel approach to estimate the precision of LoD.

1. Introduction

Arguably among the most critical performance parameters for a diagnostic procedure are those related to the minimum amount of target that can be detected and quantified [11]. The parameters describing those properties are known as the limit of detection "LoD" and the limit of quantification "LoQ". Their definitions vary slightly among regulatory bodies and standards organizations [1]. The Clinical Laboratory Standards Institute (www.clsi.org), for example, defines LoD as "the lowest amount of analyte (measurand) in a sample that can be detected with (stated) probability, although perhaps not quantified as an exact value" [2]. In many clinical laboratories and diagnostic applications, LoD is used interchangeably to "sensitivity", "analytical sensitivity" and "detection limit." This may, however, be confusing as "sensitivity" is also used in other ways. For example, in some applications "sensitivity" refers to the slope of the calibration curve, which is the definition used by the International Union of Pure and Applied Chemistry (IUPAC). CLSI defines LoQ as "the lowest amount of measurand in a sample that can be quantitatively determined with {stated} acceptable precision and stated, acceptable accuracy, under stated experimental conditions" [2]. An alternative LoQ based on clinical sensitivity and specificity has been proposed for diagnostic purposes [17].

Definitions by CLSI

LoD = the lowest amount of analyte (measurand) in a sample that can be detected with (stated) probability, although perhaps not quantified as an exact value.
LoQ = the lowest amount of measurand in a sample that can be quantitatively determined with {stated} acceptable precision and stated, acceptable accuracy, under stated experimental conditions

By far most measuring techniques generate a signal response that is proportional to the amount of measurand present. For example,

* Corresponding author. Present address: Multid Analyses AB, Odinsgatan 28, 41103 Gothenburg, Sweden.
E-mail address: amin.forootan@multid.se (A. Forootan).

http://dx.doi.org/10.1016/j.bdq.2017.04.001
Received 30 September 2016; Accepted 3 April 2017
Available online 29 April 2017

measured absorption is proportional to the concentration of the measurand as predicted by the Beer-Lambert law. Linear measurements typically generate a background signal that is observed in the absence of measurand and must be subtracted from the measured values. This background signal limits the sensitivity of the measurement and is used to estimate LoD [3]. Working at 95% confidence, the limit of blank "LoB" is:

$$LoB = mean_{blank} + 1.645 \times \sigma_{blank} \quad (1)$$

where σ is the standard deviation, and

$$LoD = LoB + 1.645 \times \sigma_{low\ concentration\ sample} \quad (2)$$

This is also the recommended estimates in the CLSI guideline EP17 [2]. The σ in Eqs. (1) and (2) refers to the true standard deviation, while SD refers to estimated standard deviation from experiments. Replacing σ for SD requires also replacing 1.645 for the corresponding t-value, which depends on the degree of freedom and, hence, the number of replicates performed).

The above equations assume response is linear and data are normal distributed in linear scale. Small deviations from normal distribution when estimating SD have been discussed [12] but in qPCR, not even the response is linear. The measured *Cq* values are proportional to the log base 2 ($\log_2$) of the concentration of the measurand (or the number of target molecules present), which is a logarithmic response. This has dramatic implications on the analysis and interpretation of the data [4]. For example, no Cq value is obtained when a negative sample is measured, as the response never reaches the threshold line, and the standard deviation (SD) cannot be calculated for any set that includes negative samples. Hence, it is not possible to estimate LoD and LoQ by the standard procedures above. A further complication is that estimating confidence intervals assumes normal distribution. While linear data often are normally distributed in linear scale, qPCR data show normal distribution in logarithmic scale, further disqualifying the conventional approaches. To estimate LoD in qPCR one needs to revert to the definition of LoD In this paper we present the procedure to experimentally determine LoD based on sample replicates and also a novel method to estimate the confidence of the LoD. We also present the procedure to estimate LoQ of a qPCR system.

2. Materials and methods

The qPCR method used as an example to assess performance was ValidPrime [5], which is a optimized probe-based assay targeting a highly conserved, non-transcribed locus present in exactly one copy per haploid human genome. The test material was human genomic DNA (CAT# CHG50, TATAA Biocenter) calibrated against the National Institute of standards and technology (NIST) Human DNA Quantitation Standard (SRM 2372). A 2-fold dilution series was prepared covering the range 1 to 2048 molecules per reaction volume. Each standard sample was analyzed in 64 replicates, except for the most diluted sample, which was analyzed in 128 replicates. Grubb's test [13] was used to identify nine outliers that were removed, leaving 759 data points for the analysis.

In the qPCR reaction, TATAA Probe GrandMaster Mix L-Rox was used and the final concentration of the ValidPrime assay in the reaction was 200 nM of a FAM-labeled probe and 400 nM of each primer. The IntelliQube*[1] (LGC Douglas Scientific) was used for all sample and master mix dispensing, thermal cycling, and real-time fluorescence detection, utilizing 1.6 µl reaction volumes. The 2-step qPCR protocol included a 1 min enzyme activation step at 95 °C, followed by 50 cycles of 10 s at 95 °C and 30 s at 60 °C. An auto-baselining method was used when plotting the amplification curves. Cq values were calculated with the IntelliQube software by manually setting a threshold line in the region of exponential amplification across all the amplification plots.

Cq data from the IntelliQube were preprocessed and analyzed using GenEx (MultiD analyses AB).

Coefficient of variation was calculated as:

$$\sqrt{exp^{(SD(\ln(conc)))^2} - 1} \quad (3)$$

assuming log-normal distribution of replicate concentrations [6]. This follows from the fact that if the stochastic variable X has lognormal distribution, then by definition $\ln(X)$ is normally distributed with, say, mean μ and standard deviation σ. The distribution function of X is then given by

$$F(x) = P(X < x) = P(\ln X < \ln x) = P\left(\frac{\ln X - \mu}{\sigma} < \frac{\ln x - \mu}{\sigma}\right) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \quad (4)$$

where Φ is the distribution function of the standardized normal distribution. The probability density function of X is readily obtained as $F'(x)$ from which the mean $e^{\mu+\sigma^2/2}$ and variance $(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$ of X are straightforwardly obtained by integration. The coefficient of variation is defined as the ratio between the standard deviation and the mean, i.e., the coefficient of variation of X becomes $\sqrt{e^{\sigma^2} - 1}$.

Cq values were measured at p different concentrations, c_i, $i = 1, \ldots, p$, with n replicas at each concentration. For simplicity of presentation n is kept constant. The data analysis below is straightforward to generalize to the case where there are different of number of replicas for different concentrations. The resulting *Cq* values are arranged in a data matrix $(Cq)_{i,j}$ $i = 1, \ldots, p$ $j = 1, \ldots, n$ and an indicator function

$$I_{i,j} = \begin{cases} 1 & if\ \ Cq_{i,j} < C_o \\ 0 & if\ \ Cq_{i,j} > C_o \end{cases} \quad (5)$$

where C_o is a user specified cut-off value. Let $z_{i=\sum_{j=1}^{n} I_{i,j}}$ be the number of detected values at concentration c_i. The logistic regression model assumes that the observed z_i is binomially distributed, $Bin(n, f_i)$, with

$$f_i = \frac{1}{1 + e^{-\beta_0 - \beta_1 x_i}} \quad (6)$$

where x_i denotes $\log_2 c_i$. The two unknown parameters β_0 and β_1 are approximated by maximum likelihood (ML) estimation. The likelihood function is

$$L = \prod_{i=1}^{p} \binom{z_i}{n} f_i^{z_i} (1 - f_i)^{n - z_i} = \left(\prod_{i=1}^{p} \binom{z_i}{n}\right) e^{\beta_0 y_1 + \beta_1 y_2 - \varphi} \quad (7)$$

where $y_1 = \sum_{i=1}^{p} z_i$ and $y_2 = \sum_{i=1}^{p} z_i x_i$ and $\varphi = \sum_{i=1}^{p} n \ln(1 + e^{\beta_0 + \beta_1 x_i})$. Setting the derivatives of L with respect to β_0 and β_1 to zero gives the system of equations for the ML estimate,

$$\begin{aligned} y_1 - \textstyle\sum_{i=1}^{p} \frac{n}{1 + e^{-\beta_0 - \beta_1 x_i}} &= 0 \\ y_2 - \textstyle\sum_{i=1}^{p} \frac{n x_i}{1 + e^{-\beta_0 - \beta_1 x_i}} &= 0 \end{aligned} \quad (9)$$

This non-linear system of equations is solved by GenEx 6 [11], using a quasi-Newton method. The ML solution will be denoted by $\hat{\beta}_0$ and $\hat{\beta}_1$. The logistic regression curve is obtained by plotting

$$\hat{t} = \frac{1}{1 + e^{-\hat{\beta}_0 - \hat{\beta}_1 x}} \quad (10)$$

versus $x = \log_2 c$. The hat-notation indicates that $\hat{t}$ is the ML estimate of the exact $t = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}}$. The observed values y_1 and y_2 can be considered as samples from a stochastic variable (Y_1, Y_2) with distribution function $\sim e^{\beta_0 y_1 + \beta_1 y_2 - \varphi}$. The moments of (Y_1, Y_2) are obtained in terms of partial derivatives of φ with respect to β_0 and β_1, by differentiating the normalization condition

[1] For research use only. The products of LGC Douglas Scientific are not for use in diagnostic procedures.