Biomolecular Detection and Quantification 10 (2016) 24–30

Contents lists available at ScienceDirect

Biomolecular Detection and Quantification

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

Research paper

Designing and interpretation of digital assays: Concentration of target in the sample and in the source of sample

Pawel R. Debski[a,b], Piotr Garstecki[a,b,*]

[a] Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka 44/52, Warsaw 01-224, Poland
[b] Curiosity Diagnostics Sp. z o.o. Kasprzaka 44/52, Warsaw 01-224, Poland

ARTICLE INFO

Article history:
Received 25 February 2016
Received in revised form 18 April 2016
Accepted 20 April 2016
Available online 17 May 2016

Keywords:
PCR
Digital techniques
Assay
DNA
Diagnostics

ABSTRACT

We explain how to design classic digital assays, comprising identical partitions, in order to obtain the required precision of the estimate within a defined range of concentrations. The design, including the number and volume of partitions, depends significantly on whether the assay is to assess the concentration of the target analyte *in the sample* or *in the source of the sample* (e.g. a patient body) with a given precision. We also show how to translate the result referring to the concentration in the sample into the concentration in the source of the sample, including the significant change in the breath of the confidence intervals.

1. Introduction

Here we discuss explicitly the significant differences in the precision of estimates provided by digital assays, depending on whether the assay addresses the concentration of the target in the sample, or in the clinical source of the sample (e.g. a human body). The two estimates enforce different statistical description of the signals in the assay and require a different design of the experimental protocol to provide the estimates with required quality.

DNA-, RNA- and immuno-diagnostics require assays that provide precise, quantitative answers and that address wide dynamic ranges of detectable concentrations. Digital assays [[1–6]; the history of digital PCR is described in a review article by Morley] introduced absolute and precise quantization. The performance of the digital analytical techniques is typically parameterized by the dynamic range of concentration within which a given standard deviation (or variance) of the estimate is guaranteed. It is not commonly appreciated that both the standard deviation of the estimate and the dynamic range of the digital assessment depend critically on the method of calculation and of the interpretation of the result. The proper understanding of the results generated by digital assays is also important in optimization [7]. Here we explain the full analytical procedure, the sources of errors and the proper understanding of the estimate of concentration of the target marker.

Standard analytical techniques use a single 'analogue' measurement on the sample to estimate the concentration of the analyte via a comparison to a calibration curve (of e.g. absorbance of light by reference samples). Digital analytical assays, first proposed by McCrady in 1915 for quantification of bacteria and later developed in 1990s for PCR by Sykes [5], and Vogelstein and Kinzler [2], split the sample into a large number of partitions, each later separately amplified for signal. The initial concentration of the target analyte is estimated from the fraction of positive end-point signals. Positive signals typically reflect the presence of at least one molecule of the target analyte. (Fig. 1).

Digital assays have a number of attractive characteristics. Most importantly, they provide absolute estimate of concentration, increasing accuracy and alleviating the need to use reference standards. Digital methods can also be highly sensitive (from 1 molecule/assay) and highly precise. These features prompt the widening use of digital techniques, especially in oncology and in the detection of very minute quantities of diagnostic markers but also progressively as standard quantitative assays for DNA markers [8]. Moreover, digital techniques, and therefore the methodology we presented here, may find use in quantitative identification of viral and microbial pathogens in physiological samples in assays pre-

Abbreviations: PCR, polymerase chain reaction; MPN, most probable number; MC, monte carlo (simulation).

* Corresponding author at: Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka 44/52, 01-224 Warsaw, Poland.
E-mail address: garst@ichf.edu.pl (P. Garstecki).

http://dx.doi.org/10.1016/j.bdq.2016.04.002

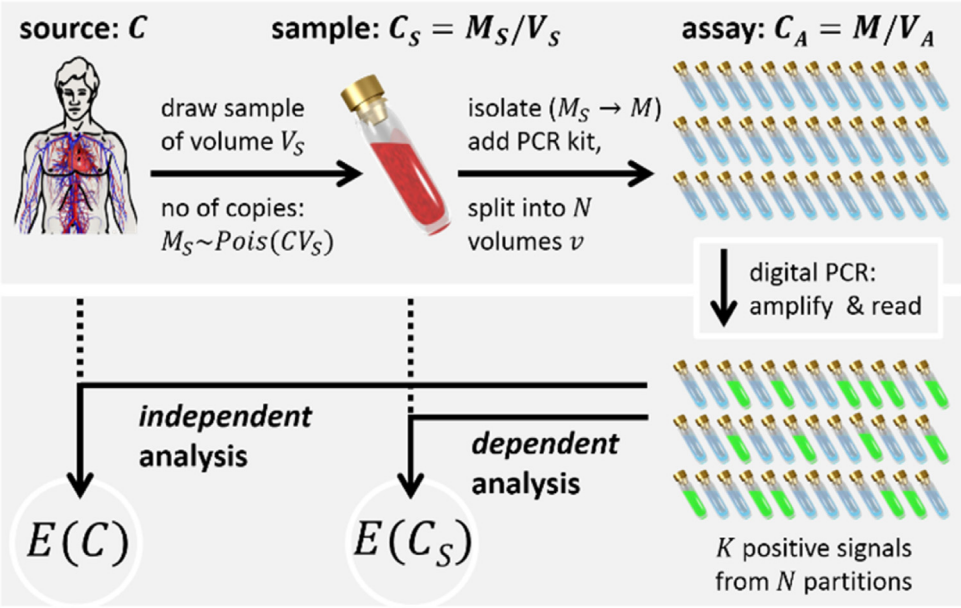

Fig. 1. The complete protocol for a digital assay. The upper shaded area shows the medical and analytical procedure. The source (e.g. a human body) presents an unknown concentration C of the analyte. A small sample of volume V_S drawn from the source contains M_S molecules. The concentration $C_S = M_S/V_S$ of target in the sample is — especially for small V_S and C— a *stochastic variable* of C. The sample then undergoes a procedure of isolation of nucleic acids and the elute (containing $M \leq M_S$ molecules) is mixed with reagents for PCR to a final volume of the assay V_A. The PCR-ready mixture is then divided into N partitions, with the molecules of the target randomly distributed between them. Amplification via PCR yields the raw result: K positive signals from N partitions. This result can be then translated via analysis (bottom shaded area) – either within a *'dependent'* scheme to yield $E_D(C_A)$ corresponding to the estimate $E(C_S)$ of concentration in the sample, or, via the *'independent'* analysis to yield $E_I(C_A)$ corresponding to the estimate $E(C)$ of the concentration of target in the original source of sample. Each of these estimates has a different confidence interval. The schematic picture of the human body is based on graphics available at: http://cliparts.co/, and https://commons.wikimedia.org/wiki/ File:Vein_art_near.png.

pared for the point-of-care formats. Looking further, finding new chemically specific chain- and avalanche reactions that could serve for amplification of the presence of small counts of molecules of analyte could become an important direction in the development of new analytical methods in chemistry, biochemistry and medicine [9].

The analytical procedure usually focuses on determining the concentration C of the target analyte in the source (e.q. human body). It begins with drawing a sample of volume V_S that contains a stochastic number M_S of molecules of the target. The number of molecules is a Poisson random variable with average value equal $C \times V_S$. Therefore, $M_S \sim Pois(CV_S) = \frac{\exp(-CV_S)\times(CV_S)^{M_S}}{M_S!}$. The concentration C_S in the sample is thus a stochastic variable of C and of V_S: $C_S = M_S/V_S \sim \frac{1}{V_S} \times Pois(CV_S) = \frac{1}{V_S} \times \frac{\exp(-CV_S)\times(CV_S)^{M_S}}{M_S!}$. The sample is then purified and mixed with the reagents for the PCR assay. This results in a volume of the assay $V_A = V_{elution} + V_{PCR}$. The volume of the assay is then split into N compartments, each of volume v, $Nv = V_A$. After the PCR reaction on all of the individual compartments, we read the raw result, i.e. K positive signals from N partitions. This raw result can be then translated into the estimate of the initial concentration of the analyte in the assay $E(C_A)$ that is provided with a finite relative standard deviation $\sigma(C_A)$, which can also be treated as the precision of the estimate.

As we show below, the analysis can be performed via two schemes. In the first, one assumes that the probability of finding a molecule of target in any partition depends on the placement of target analyte in all other compartments in the assay. This corresponds to the actual situation of drawing the molecules from a finite pool containing only M target molecules. In this context, the finite pool is the volume of an assay. This (*dependent*) scheme can be translated into the estimate of the concentration of target in the sample: $E(C_S) = \alpha E_D(C_A)$, and $\sigma(C_S) = \sigma_D(C_A)$, where α is a numerical factor reflecting the change in the volume between the sample and the assay and the efficiency of isolation $\eta = M/M_S$: $\alpha = (V_A/V_S)/\eta$. However, this estimate is limited only to the volume of the sample, which is usually very small compared to its source, i.e. human body. Therefore, the precision of the estimate provided by the test, which is typically given by the producer of a dPCR system, might be misleading if the concentration in the human body is of interest.

In the second (*independent*) scheme, the estimate is provided via an analysis based on the superficial assumption that the probability of finding a molecule of target in any particular partition is independent of all the other ones (just as if the material for each partition was drawn directly from large reservoir – a source of the sample, i.e. human body). Via this analysis we can obtain the estimate of the concentration of the target in the

Source: $E(C) = \alpha E_I(C_A)$, and $\sigma(C) = \sigma_I(C_A)$. Therefore, the precision of the estimate, calculated in this approach, provided by the test concerns the concentration in the human body.

Below we show how the two methods of analysis of digital estimates can be conducted, how different are the relative standard deviations σ_D and σ_I, and how to design digital assays for the required precision of assessment of the concentration in the sample or in the source over a required range of concentrations.

2. Results

In a classic digital assay [2,10,11] the sample is equally divided between a number N of identical partitions. All the partitions are then amplified via PCR after which the binary signal is read from each partition. Once the signals are collected, they must be translated into the estimate of the (a priori unknown) concentration of the analyte in the sample, and concentration of the analyte in the source of the sample (e.g. in a human body). (Fig. 2).

Current mathematical procedures used to analyze the outcome of the assay are based either on the Most Probable Number (MPN)