Statistical Estimations of PCR Amplification Rates

Jean Peccoud and Christine Jacob

INTRODUCTION

Quantitative applications of the Polymerase Chain Reaction (PCR), also known as Quantitative-PCR (Q-PCR) are intended either to determine the number of copies of a given nucleic acid sequence, or more generally, to determine the relative abundance of two sequences. Current methods to determine exact numbers of molecules overcome the determination of the amplification rate by assuming identical amplification rates for a target DNA sequence and a standard of known quantity introduced into the experiment design, so that only the ratio of amplified products need be determined. Violations of the hypothesis of identical amplification rates for two sequences will result in a systematic bias in the experiment results that underestimates or overestimates the initial copy numbers. Acquisition of kinetic PCR data was pioneered by Higuchi *et al.* (Higuchi et al., 1993; Higuchi et al., 1992) and commercial instruments have been available since early 1996. Kinetic data provide a new way to determine the amplification rate, and we can foresee that their availability will rekindle interest in the algorithms used to compute the initial quantities of DNA sequences. Analysis of kinetic PCR patterns will soon make its way into the family of recipes that have been in use for some years in this field. This chapter provides evidence that a statistical analysis of the amplification rate is critical to ensuring a reliable estimate of the initial copy number.

PCR AMPLIFICATION LEADS TO STOCHASTIC FLUCTUATIONS.

PCR is an exponential amplification of a DNA target molecule population of initial quantity, N_0. If every molecule were duplicated at each cycle, the population size at cycle n, N_n, would then be twice the size of the population at

cycle *n*-1. Following this reasoning leads to the formula $N_n = 2^n N_0$ which would be a convenient basis for the derivation of the initial copy number from the size of the population after *n* cycles of amplification.

$$N_0 = \frac{N_n}{2^n} = 2^{-n} N_n$$

Unfortunately, things are not that simple. The yield of the amplification reaction is not 100% and thus, the amplification rate *m* is less than 2, and in practice 1<m<2. As we shall see, the surprising consequence is that the behavior of the PCR reaction is no longer deterministic. When the yield of the reaction is 100%, there is a non ambiguous relation between the initial copy number and the number of molecules after *n* cycles of amplification. As soon as it is possible to measure the number of the amplification products, one can extrapolate to the initial number of molecules that were amplified. The only other deterministic case occurs when the amplification rate is 1, which means that no amplification occurs at all, but this case has no more practical interest than the previous one.

When the amplification rate is between 1 and 2 - the usual case - then reaction dynamics become stochastic, as the following discussion will show. The reaction yield $r = m$-1 describes a probability that a given molecule will be duplicated during one cycle of amplification. Suppose that a single initial molecule undergoes PCR with a reaction yield *r*; then, after one cycle of amplification, the number of molecules can either be 2 with a probability *r*, or remains 1 with a probability 1-*r*. As the reaction proceeds, the number of molecules after cycle *n* will be randomly distributed between 1 and 2^n. Suppose now that instead of having a single initial molecule, there were 2 initial molecules. Then, the number of molecules after *n* cycles in the reaction is randomly distributed between 2 and 2×2^n. Consider the consequence of this with respect to determining the initial copy number: for any PCR that results in a number of molecules N_n greater than 2 but less than 2^n, it is no longer possible to determine with certainty whether the initial copy number was 1 or 2 molecules. Some information was lost during the course of the amplification.

It is worth stressing that this argument is not based on any measurement error in the amplification rate or in the number of amplification products at cycle *n*, and introducing such errors would make the determination of the initial quantity even more challenging.

It is the random or stochastic behavior of PCR itself that requires a suitable statistical analysis. Initial copy numbers cannot be determined, they can only be statistically estimated. This means for instance that the relation below is not rigorously correct:

$$N_0 = \frac{N_n}{m^n}$$

Instead, it holds approximately:

$$N_0 \quad \frac{N_n}{m^n} \tag{1}$$

The meaning of the sign can be made more precise:

1. The larger N_0 is, the better is the approximation of N_0 by $N_n m^{-n}$:
   at any cycle *n*, $\lim_{N_0 \to \infty} \frac{N_n m^{-n}}{N_0} = 1$

2. The mean value $E_{N0}\left(N_n m^{-n}\right)$ of the $N_n m^{-n}$ obtained from independent replicate amplifications each starting from N_0 molecules and each having the same amplification rate *m* is equal to N_0 whatever the value of N_0 :
   for any N_0 , $E_{N0}\left(N_n m^{-n}\right) = N_0$

3. When *n* is large enough, $N_n m^{-n}$ is approximately equal to a random variable $W_{N_0,m}$ with expectation (mean value) N_0 and variance $(2-m)m^{-1}$:
   $\lim_{n \to \infty} \frac{N_n}{m^n} = W_{N_0,m}$ with $\mathrm{E}\left(W_{N_0,m}\right) = N_0, \sigma^2\left(W_{N_0,m}\right) = \frac{2-m}{m} N_0$

Although these properties may seem a bit technical at first glance, they are a very good expression of how far an estimation of the initial copy number based on (1) might be from the actual value of N_0 . The ratio of the standard deviation of the estimate over its mean value is a simple indicator of the dispersion of this estimation. In practice, kinetic PCR experiments usually involve enough cycles prior to observation so that this limit property can be used in practice. If 1000 molecules are amplified with an amplification rate of 1.80, then the dispersion of the estimation of the initial copy number based on (1) is:

$$\frac{\sqrt{\frac{2-m}{m} N_0}}{N_0} = \sqrt{\frac{2-m}{mN_0}} = \sqrt{\frac{2-1.80}{1.80 \times 1000}} = 1\%$$

This computation demonstrates that for initial copy number numbers greater than 1,000, (1) allows them to be estimated rather precisely. However, for initial copy numbers less than 1,000 the precision of the estimation may become limiting and should then be provided with a confidence interval for the actual initial copy number. For instance, the estimation of the initial copy number computed from an amplification starting with a single molecule and with a rate of amplification 1.80, has a relative dispersion equal to $\sqrt{0.2/1.80} = 33\%$. Actually, the relative uncertainty of the quantitative measurement based on a PCR amplification can be derived from the formula used to compute the confidence interval of the initial copy number estimation. These results demonstrate that for low copy numbers, the measurement uncertainty is significantly greater than the dispersion indicator computed here. Uncertainties range from 100% for a few copies to 10-20%, depending on the amplification rate, for initial copy numbers close to 100. (Peccoud and Jacob, 1996). Several authors have reported on the difficulty of obtaining reproducible amplification results when starting from low copy numbers (Lantz and Bendelac, 1994; Piatak et al., 1993; Karrer et al., 1995). Apart from the inherent sampling errors that result from the manipulation of such low numbers of molecules, the inherent stochastic fluctuations of PCR dynamics itself may explain a large part of the dispersion in their results.

ESTIMATIONS OF INITIAL COPY NUMBERS MUST BE BASED ON AMPLIFICATION RATE ESTIMATIONS

Since most research need to quantitate copy numbers below 1,000, this section focuses on amplifications starting with high copy numbers (>1000). It will then be considered that:

$$N_0 = \frac{N_n}{m^n} \tag{2}$$

Even in this restricted perspective, the estimation of the amplification rate could be extremely valuable. Since kinetic data have not been previously available, rigorous estimation of the amplification rate was difficult, so methods have been developed to bypass this step in the analysis of amplification data. Most of these techniques rely on standard sequences introduced in known quantities into the experiment design. Assuming that the standard and the target sequence (the one that must be quantified) have identical amplification rates, one can determine the initial copy number of the target, $N_{n,T}$, from the initial quantity of the standard, $N_{0,S}$, and the measurements of the amount of two amplified sequences, $N_{0,S}$ and

$N_{n,T}$, without any direct computation of the amplification rate. The basis of this approach is the next relation which can easily derived from (2):

$$\frac{N_{n,S}}{N_{n,T}} = \frac{N_{0,S}}{N_{0,T}} \left(\frac{m_S}{m_T}\right)^n$$

Given the hypothesis that $m_S = m_T$, the ratio of the standard molecule number over the number of target molecules remain constant after any number of amplification cycles. However, violations of this hypothesis will result in a significant evolution of the ratio over time, changing as $\left(m_S m_T^{-1}\right)^n$. For instance, when $m_S = 1.9$, $m_T = 1.8$, and n=25, then there is a 3.86 fold difference between $N_{n,S}/N_{n,T}$ and $N_{0,S}/N_{0,T}$. Factors causing different amplification rates for the standard and the target are probably more numerous than factors causing exactly the same amplification rates. Disparity may arise from minor differences in the sequences, from tube to tube differences, sample to sample differences, and so on. Since the methodology based on ratios arose at a time when the amplification rates were difficult to measure, it is likely that violations of the hypothesis would not be detected and taken into account, and quantitative estimations would then be contaminated with a systematic error.

When Q-PCR experiments are conducted for relative quantification purposes such as the comparison of the quantities of two molecules, a common situation in gene expression and mRNA quantification experiments, the same argument applies. Quantitative differences between the two sequences might be underestimated or exaggerated by small variations in their respective amplification rates.

THE AMPLIFICATION RATE ESTIMATORS

The previous section emphasized the need for a method to estimate the amplification rate. When examined carefully, the requirements for such a method are very stringent.

1. The estimation of the amplification rate of a reaction must be based on the data collected from this reaction only. It cannot be based on a set of related reactions since there are variations of the amplification rate from one reaction to another.

2. The estimation must be able to detect the end of the so called "exponential phase", the early phase of the reaction during which the amplification sustains a steady rate of amplification.

3. The estimation must be computed from the measurement of the DNA molecule numbers and not from the molecule numbers themselves.

Requirement 3 may appear naive but is important since real-world measurements are always contaminated with errors. The most common model of a measurement assumes that a measure, X_n, is proportional to the measurand, here the number of molecules N_n, plus a random value ε_n with a constant statistical distribution that is usually assumed to be Gaussian, $N(0,\sigma^2)$.

$$X_n = a\ N_n + \varepsilon_n \tag{3}$$

A convenient way to meet requirement 1 is to use a set of kinetic data collected during a single amplification. Requirement 2 will be met only if the estimator can be computed on a subset of the kinetic data and if it is sensitive enough to detect the decay of the amplification rate in data collected during the late phase of the reaction.

We have previously proposed and characterized the estimator $\hat{m}_n$ (Jacob and Peccoud, 1996a; Jacob and Peccoud, 1996b) which fits the 3 requirements aforementioned:

$$\hat{m}_n = \frac{X_{n-2} + X_{n-1} + X_n}{X_{n-3} + X_{n-2} + X_{n-1}} \tag{4}$$

As *n* grows, this estimator converges exponentially towards the actual value of the amplification rate, $\lim_{n\to\infty} \hat{m}_n = m$. In principle, the speed of convergence depends on two parameters: the amplification rate and the initial copy number. However, simulations of kinetic PCR data with increasing errors demonstrate that the measurement error can significantly delay the observation of this convergence (see Fig. 3). Another interesting use of $\hat{m}_n$ is the ability to monitor on a cycle by cycle basis, the evolution of the amplification rate during the late phase of the reaction. In this case, the estimator is very close to the actual value of *m* since the number of cycles is high enough. This possibility is particularly valuable to detect the end of the exponential phase.

In order to avoid confusion that may arise from a previous paper (Peccoud and Jacob, 1996), it is worth noting that $\hat{m}_n$ is not the only possible estimator. There is in fact an entire family of valid estimators that have identical limit properties, though they may differ with respect to their sensitivity to measurement errors. For technical reasons beyond the scope of this paper, only the estimator (4) will be considered here. Finding the most suitable estimator to use on noisy data is still an interesting field of investigation.

Before proceeding to an estimation of amplification rates on real data, another tool is still required to measure the convergence of the estimator. Graphical representations of the estimations will provide a visual appreciation of their convergence, but an index is needed to quantify the convergence, and one natural index is based on the successive differences of the estimations:

$$\delta_n = \sqrt{\left(\hat{m}_n - \hat{m}_{n-1}\right)^2} = \left|\hat{m}_n - \hat{m}_{n-1}\right|$$

Here δ_n is the absolute value of the difference between two successive estimations. The lower its value is, the closer are two successive estimations of the amplification rate.

DATA SET

The ABI PRISM™ 7700 Sequence Detector system by PE Applied Biosystems is the first commercially available instrument to produce kinetic data of PCR amplifications. A data set representative of the performance of this instrument was kindly provided to us by Ray Lefebvre and Lincoln McBride of Perkin-Elmer.

A detailed description of the 7700 system can be found elsewhere in this book ****(give ref)****. At installation, the Install Kit is run to validate instrument performance. The details of the protocol used to amplify the kit are provided in the instrument manual. For our purpose, only the general structure of the experiment is of interest (Table 1). The No Template Control (NTC) is a reaction conducted in normal conditions except that the DNA template solution is replaced by TE buffer. The last two series of replicates (10,000 and 5,000) are treated as unknowns. Their initial copy numbers are derived from the standard curve generated from the standards, and compared to their known copy numbers to validate the instrument performance.

The data that will be analyzed in this paper are called the "clipped data" by Perkin Elmer, which is the normalized fluorescence from the reporter dye at the end of each extension phase (FAM dye was the reporter in this case).

Wells	**Initial Copy Number**
1-4	No Template Control
5-8	1,000 copies standard
9-12	2,000 copies standard
13-16	5,000 copies standard
17-20	10,000 copies standard
21-24	20,000 copies standard
25-60	10,000 copies
61-96	5,000 copies

Table 1 The 7700 Installation Kit plate layout.

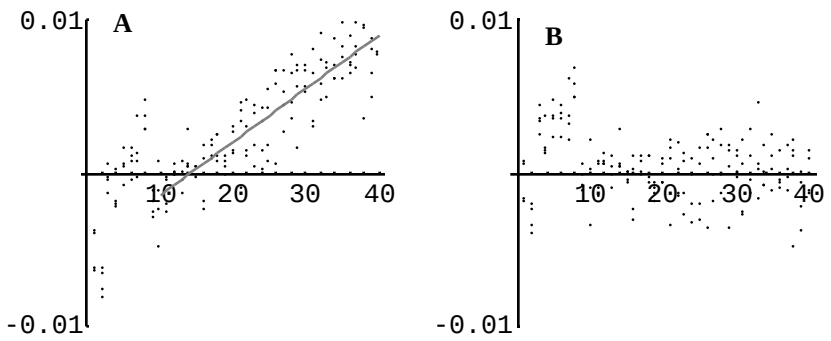

FIGURE 1 Correction of the NTC trend: The No Template Control data are pooled and plotted on A. There is a significant trend that is materialized by the gray line. On B, the data were corrected for the trend. All other data were corrected similarly.

ANALYSIS OF THE NO TEMPLATE CONTROL AND CORRECTION FOR THE BACKGROUND TREND

Before proceeding to an estimation of the amplification rate, it is necessary to ascertain as much as possible that the data fit the model (3). One way to do this is to carefully analyze the data of the NTC replicates. Since no reaction occurs in these wells, their fluorescence should not increase during the 40 cycles of amplification; they should only be subject to random fluctuations that result from measurement errors of the instrument.

When the NTC data are plotted together (the four replicates are pooled into single data set), it appears that their fluorescence has a very significant growth trend which can be observed in Fig. 1A. The origin of the growth of the background fluorescence is difficult to figure out. A non-enzymatic degradation of the TaqMan probe may occur during the thermo-cycling as a result of the incubation at high temperature, or the laser illumination used to excite the fluorescent dyes may gradually break some of the bonds of the probe over time. This phenomenon does not seem to be documented so far. The growth of the background fluorescence is not completely linear; there is a peak between cycles 1 and 10, afterwards the growth is more regular. Since the data collected before cycle 10 are extremely noisy and thus useless in the computation of the initial copy number, it is not necessary to estimate precisely this peak and these points

can be set aside.The background fluorescence measured between cycle 10 and cycle 40 is fitted to a linear model that can be regarded as the average growth of background fluorescence, shown in gray in Fig. 1A. The linear model can be used to refine the data, and corrected background fluorescence is plotted in Fig. 1B. The correction is effective since a trend is no longer visible. This can be confirmed statistically by computing the mean value of the refined data collected after cycle 10 which is equal to $-9.7\ 10^{-18}$. This is not significantly different from 0. The standard deviation of this subset of the refined background fluorescence is $1.66\ 10^{-3}$, a value which can be regarded as the standard deviation of the instrument measurement error, a point which will be addressed in more detail below. All the data are corrected in a similar way and only this refined data set will be discussed in the following sections of this chapter.

ANALYSIS OF ONE AMPLIFICATION REACTION

Let us now apply the tools that were introduced in the previous sections. One reaction, well 7, was chosen to construct Fig. 2. The series of $\hat{m}_n$ is first computed. Note that this computation is very simple to implement and could readily be programmed in a spreadsheet application.

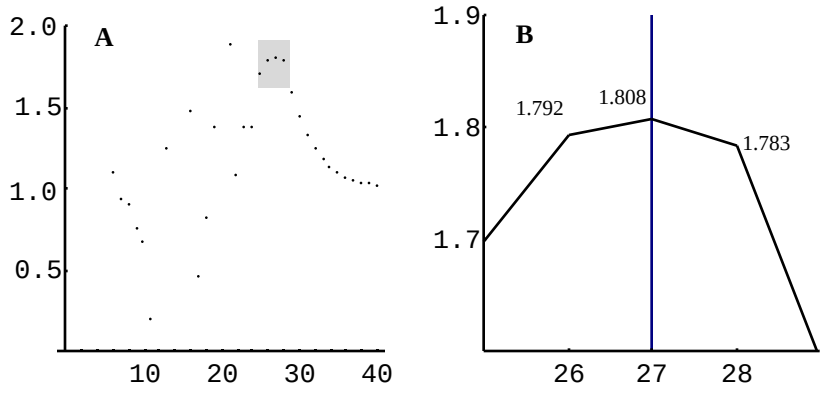

FIGURE 2 Analysis of a typical amplification: The amplification rate is estimated from the well 7. Part B is a zoom on the gray area of part A.

The general behavior of the series can be understood from its graphical representation in Fig 2A. Three phases can be distinguished. The first phase

extends from cycle 1 to cycle 22; in this range the estimator behavior is extremely erratic, jumping from very low negative values (-4.74 at cycle 4) to very high positive values (2.7 at cycle 20). During this phase, the signal does not rise significantly above the background, and the large fluctuations of the estimator result from the ratios of relatively small numbers. Since the noise fluctuates around 0, the sum of three successive measurements can be either negative or positive, thus the sign of $\hat{m}_n$ changes frequently. The mean value of the estimator in this phase is 0.95 indicating that despite the large fluctuations, the average amplification rate estimation is close to 1. At this stage, then, no amplification has yet been detected.

The second phase extends from cycle 23 to cycle 27, corresponding to the rising slope of the of the estimator curve that reaches its maximum at cycle 27. The signal itself starts to rise above the background noise, along with the amplification rate estimator. But since noise is still a large fraction of the signal, the rate estimator grows toward its limit value. Due to the noisy component of the measure, the growth is not always very smooth.

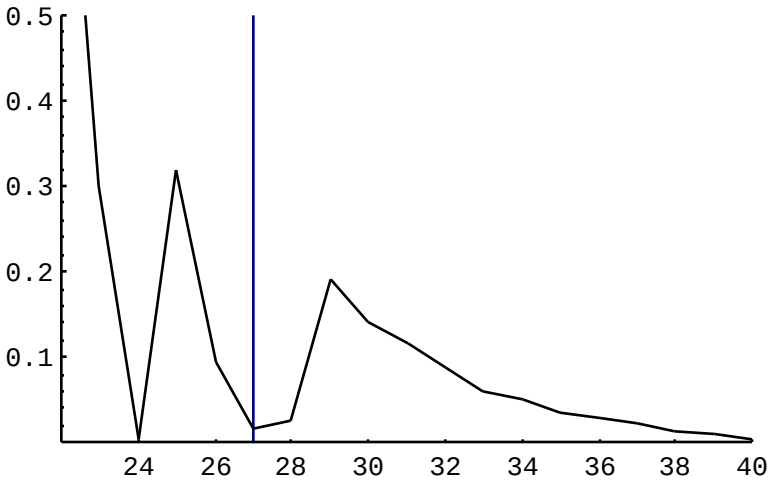

FIGURE 3 Analysis of the $\hat{m}_n$ convergence: The convergence index δ_n minimum at cycle 27 which is the end of the exponential phase. The minimum at cycle 24 is ascribed to random fluctuations.

The third and last phase of the reaction appears as an exponential decay of the estimator values, resulting from the end of the exponential phase of the reaction. Since the real amplification rate has started to decrease, its estimator shows a similar evolution as we would expect. When the ascending and descending slopes of the peak are compared, the latter is clearly much smoother than the former since, at this time, the estimator has reached its asymptotic behavior and

the noise is no longer large enough to seriously perturb the estimator behavior, at least on the scale of this plot.

Why was the end of the exponential phase set to cycle 27? This is not because the peak reaches its maximum at cycle 27, but rather because it seems that the estimator most closely approaches its limit at cycle 27. If the amplification rate were constant during 40 cycles, then estimator fluctuations would decrease so much that they would allow a very precise determination of the amplification rate value (see Fig. 4 and the next section). On this particular data set, the situation is more complex. There are very few cycles in the exponential phase where the noise is already negligible. Prior to this phase the estimator is erratic, and afterwards it simply follows the decay of the amplification rate. Plotting the evolution of δ_n helps one to determine the end of the exponential phase. The convergence of $\hat{m}_n$ results in a local minimum for δ_n and there is actually such a minimum at cycle 27 (see Fig. 3). The problem is that there is also a minimum at cycle 24 which even lower than the one at cycle 27. Why would we not consider cycle 24 as the end of the exponential phase? From Fig. 2A, one can see that the small plateau at cycle 24 is more likely due to a random fluctuation than to the convergence of the estimator because it is too far from the peak maximum. Instead of using the δ_n plot, the highly magnified zoom on the top of the peak (gray rectangle on Fig. 2A) shown in Fig 2.B can be used to reach a similar conclusion.

It is possible to apply this method to analyze all the data collected for the standards in wells 5 to 24. The results are presented in Table 2. When it was not possible to find a cycle n such that $\delta_n < 0.05$, the result is reported as being Not Determined (ND), meaning that the convergence of the estimator was too perturbed by the measurement error to be reliably observed. When it is possible to determine, the estimation of the end of the exponential phase is reproducible and consistent with the initial copy number. There is a difference of about 4 or 5 cycles between the 20,000 copy reactions and the reactions starting from 1,000 copies. The well-to-well variation of the amplification rate estimation is less than ±0.04 which compares well with the δ_n values so that at this stage of the analysis, one cannot evaluate a possible well-to-well difference in the amplification rate. Instead, since the mean value of the amplification rate estimation is 1.855, the differences in the estimations of the end of exponential phase are explained reasonably well by the 20-fold dilution factor since $1.855^4=11.84$ and $1.855^5=21.96$.

VALIDATION OF THE MODEL AND FLUORESCENCE CALIBRATION

In order to confirm the validity of the model used to build the statistical estimators, it is worth trying to compare the analysis of experimental data with a corresponding analysis of simulated data. Data from well 22 are used for the comparison since they converge well and thus allow a precise estimate of the amplification rate. In this section, data will be simulated with the addition of an increasing level of noise until the analysis of simulated data matches well with the pattern observed in experimental data.

Well	**Cycle**	$\hat{m}_n$	δ_n
5	28	1.97	0.028
6	ND		
7	27	1.81	0.016
8	27	1.79	0.021
9	ND		
10	27	1.92	0.30
11	28	1.78	0.31
12	ND		
13	ND		
14	26	1.88	0.010
15	26	1.85	0.032
16	26	1.83	0.028
17	25	1.86	0.042
18	24	1.89	0.023
19	ND		
20	23	1.88	0.009
21	23	1.85	0.030
22	24	1.83	0.001
23	24	1.79	0.028
24	23	1.87	0.003

TABLE 2 Amplification rate estimations for wells 5 to 24.

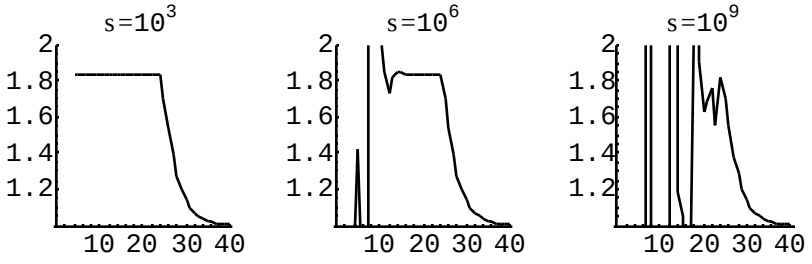

FIGURE 4 Effect of the noise level on the convergence of the amplification rate estimator. Analysis of three sets of simulated data. The parameters of the simulation

The computation of the $\hat{m}_n$ allows one to determine that the end of the exponential phase is cycle 24 where the amplification rate estimation is 1.83. For this trajectory, the initial copy number is 20,000. where the amplification rate is estimated at 1.83. For this trajectory, the initial copy number is 20,000. The only unknown parameter is the standard deviation of the measurement error. Noise of increasing levels over a range of several orders of magnitude was successively added to simulated data. The effect of the noise on the convergence of the estimator can be observed on Fig. 4. As anticipated, when the noise level becomes too large, it interferes with the ability of the estimator to converge before the end of the exponential phase. For noise with standard deviation of 4 10^8 molecules, the analysis of the simulated data and the analysis of the experimental data look rather similar. Of course, this appreciation is mainly visual (Fig. 5). Since the calibration function of the 7700 is not available, data produced by the simulation algorithm were expressed in molecule numbers. The standard deviation of the measurement error used to construct Fig. 5 is also expressed in molecule number. How can these simulated data be converted into the corresponding fluorescence units? Since the initial copy number is high, the fluctuations due to noise are limited. Thus, faithful simulation parameters should ensure that at the end of the exponential phase, i.e. at cycle 24, the data of the simulated amplification, 3.98 10^{10} molecules, and the data measured on well 22, 0.168 fluorescence units, should be approximately equal but expressed on different scales. The ratio of these two values suggests a scaling factor equal to 4.22 10^{-12} fluorescence units/molecule. We can check that this ratio is consistent with other features of our model that have been estimated on different data since it is also possible to compare the two expressions of the standard deviation of the measurement error

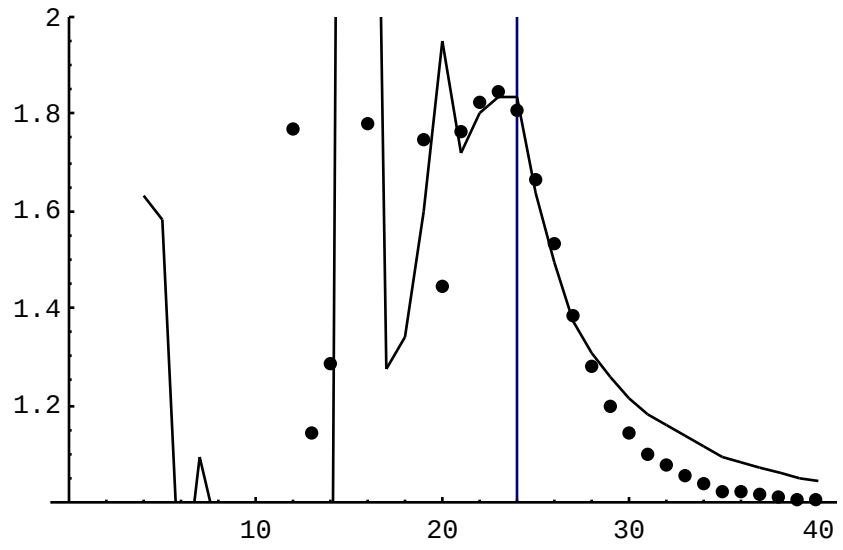

FIGURE 5 Comparison of simulated data with a 4 10^8 molecule noise standard deviation with the analysis of the data collected from well 22. For this level of noise, it is possible to obtain analysis patterns that look similar except in the late phase of the reaction. This comparison is a validation of the model used to compute the amplification rate estimations.

In order to adjust the analysis of simulated data to match the analysis pattern of experimental data, it is necessary to introduce a measurement error of 4 10^8 molecules. But the analysis of the NTC data provides another estimation of the error standard deviation: 1.66 10^{-3} fluorescence units. The product of the scaling factor and the error standard deviation, expressed in molecules, should be approximately equal to 1.66 10^{-3}:

$$4.22\ 10^{-12} \times 4\ 10^{8} = 1.68\ 10^{-3}\ .$$

Although the two numbers do not match exactly, they do look very consistent with one another. In summary then, two independent lines of reasoning and computation give consistent results and tend to demonstrate the validity of the model upon which the analysis is based.

ANALYSIS OF THE AMPLIFICATION RATE DECAY

In the simulation algorithm used in the previous section, the decay of the amplification rate is assumed to be exponential. The hypothesis is that the amplification rate looses a few percent every cycle after the end of the exponential phase. If *ne* is the last cycle in the exponential phase, and if the rate decreases by τ% per cycle, then for any *n*>*ne*, the amplification rate used in the simulation is:

$$m_n = m\ (1-\tau)^{n-ne}$$

It was puzzling that this type of decay did not permit us to match the late phase of the simulations with the experimental data. This is visible in Fig. 5 where it is apparent that there is a discrepancy between the two plots in the late phase of the reaction. This was confirmed when $\log(\hat{m}_n)$ was plotted as a function of *n*. It became obvious that the relationship between these two quantities was not linear. Since it still seemed to be decreasing very rapidly, we tried plotting $\log(\log(\hat{m}_n))$ against *n* and in this case a linear relationship finally appeared. The slope of this line is on the order of -0.18. Analysis of several trajectories gives similar results. This fit means that the time evolution of the amplification rate in the late phase of the reaction can be represented by:

$$m_n = m\ \exp(\exp(-\tau(n-ne)))$$

This is a spectacular decay. The end of the exponential phase is not marked by a slightly decreasing amplification rate but rather by a total collapse of the reaction yield. It is will be interesting to see what a possible biochemical or biophysical explanation of this observation might be.

PERSPECTIVE

The analysis of the data collected on only a few wells is presented in this paper. Some of the wells (5 out of 16) did not show a convergence strong enough to allow one to reliably determine the end of the exponential phase or the amplification rate. Moreover, at this stage of our research there is still a need for an objective criterium that could be used to completely automate the analysis of kinetic PCR data. The measurement error strongly affects the convergence behavior of the amplification rate estimator. In many cases it seems that the amplification rate starts to decrease before the signal to noise ratio increases enough to reliably determine the amplification rate.

The relative measurement uncertainty of the 7700 can be computed and is approximately equal to 1% at the end of the exponential phase. For a DNA

quantification protocol this precision is extremely good. However, it still limits the ability to determine the amplification rate with the greater accuracy needed for a precise estimation of the initial copy number. Another possibility currently being investigated is the use of estimators of the amplification rate that are less sensitive to measurement errors. In the near future, we expect that a combination of more effective estimators applied on data of higher quality will allow one to determine the amplification rate with higher precision.

Though it is still in its infancy, the statistical estimation of the amplification rate provides valuable results. The analysis of the reactions can be performed without any standard. Analysis of several replicates of identical reactions (wells 25-60 and 61-96) tends to indicate that there may be significant well-to-well variability in the amplification rate (data not shown) even though experimental conditions are identical. If confirmed, this result would indicate that the reaction is sensitive to parameters that are not controlled by current experimental setup. This would be another strong argument in favor of individual amplification rate estimations. In the most favorable reactions, it is already possible to determine the amplification rate with a 10^{-2} accuracy. As soon as the data measured by the kinetic PCR instrument can be calibrated, these amplification rate estimations could be used for an absolute quantification of the initial copy number. Comparison of the analysis of NTC data with simulated data might produce an original and powerful way to calibrate the instrument.

Finally, whatever the current limits of the analysis of kinetic PCR data as presented in this paper, it has a very nice quality. It is self validating and does not rely upon assumptions or hypothesis that cannot be verified. The only hypotheses used are qualitative and related to the dynamics of the Polymerase Chain Reaction and to the measurement of DNA molecule numbers. The comparison of simulated data with experimental data demonstrate their validity. Measurement errors are still limiting the accuracy of the amplification rate estimation. It is possible to compute a convergence index that reflects the quality of the estimates. High values of this index mean that the data are too noisy to be analyzed accurately. This convergence index should provide a solid basis to compare the quality of the data collected on different kinetic PCR machines.

REFERENCES

Higuchi R, Dollinger G, Walsh P S and Griffith R (1992): Simultaneous amplification and dection of specific DNA sequences. *Bio/Technology* 10:413-417

Higuchi R, Fockler C, Dollinger G and Watson R (1993): Kinetic PCR analysis : real-time monitoring of DNA amplification reactions. *Bio/Technology* 11:1026-1030

Jacob C and Peccoud J (1996a): Estimation of the offspring mean for a supercritical branching process from partial and migrating observations. *C R Acad Sci Paris Série I* 322:736-768

Jacob C and Peccoud J (1996b): Inference on the initial size of a supercritical branching process from partial and migrating observations. *C R Acad Sci Paris Série I* 322:875-880

Karrer E E, Lincoln J E, Hogenhout S, Bennett A B, Bostock R M, Martineau B, Lucas W J, Gilchrist D G and Alexander D (1995): In situ isolation of mRNA from individual plant cells: Creation of cell-specific cDNA libraries. *Proc Natl Acad Sci U S A* 92:3814-3818

Lantz O and Bendelac A (1994): An invariant T cell receptor alpha chain is used by a unique subset of major histocompatibility complex class I-specific CD4+ and CD4-CD8- T cells in mice and humans. *J Exp Med* 180:1097-1106

Peccoud J and Jacob C (1996): Theoretical uncertainty of measurements using quantitative polymerase chain reaction. *Biophys J* 71:101-108

Piatak M, Saag M S, Yang L C, Clark S J, Kappes J C, Luk K C, Hahn B H, Shaw G M and Lifson J D (1993): High Levels of HIV-1 in Plasma During All Stages of Infection Determined by Competitive PCR. *Science* 259:1749-1754