Quantitative polymerase chain reaction

Introduction

Quantitative polymerase chain reaction (PCR), also known as real-time PCR, is a technique used to quantify specific nucleic acid sequences in a sample of interest. The assay relies on the generation of a fluorescent signal that is directly proportional to the amount of target DNA and is used most commonly for the analysis of gene expression. This article outlines the basic principles and applications of quantitative PCR, and addresses some common pitfalls and ways they may be overcome. While this is by no means an exhaustive overview, it is hoped that this article will provide a useful introduction for the clinician beginning a period of training in basic science.

Principles of the assay

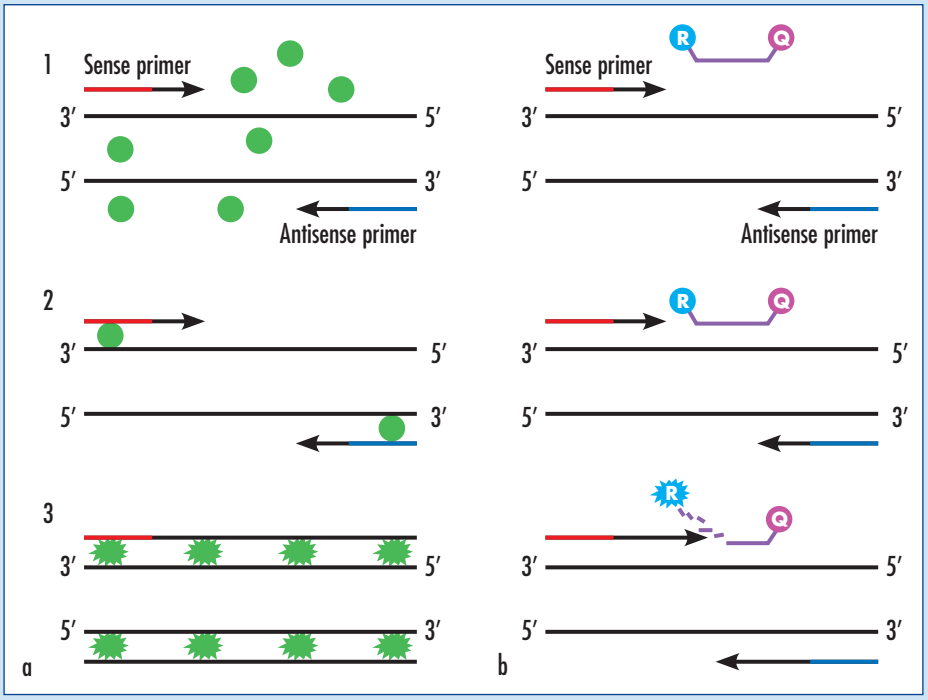
Quantitative PCR relies on the coupling of DNA amplification by PCR to a fluorescent signal that increases in intensity as more DNA is synthesized. PCR amplification relies on repeated heating and cooling, or thermal cycling, of a reaction mixture containing template DNA, sequence-specific oligonucleotide primers, nucleotides and a heat-stable DNA polymerase enzyme, such as Taq, to double a DNA sequence of interest with each cycle. The inclusion of a DNA-binding dye generates a fluorescent signal that is directly proportional to the amount of amplified target DNA, allowing accurate quantification of a sequence of interest. The detection of DNA as it is amplified contrasts with earlier techniques such as Northern blotting.

A range of fluorescent markers can be used for quantitative PCR, including DNA binding dyes such as SYBR green that intercalate with all double-stranded DNA, or sequence-specific fluorescent probes that only bind to a particular target sequence (*Figure 1*). Double-stranded DNA binding dyes have the advantage of being compatible with multiple primer sets and are often substantially cheaper than acquiring specific probes for an equivalent number of gene targets. However, the specificity required for accurate quantification is reliant solely on the sequence-specific primers, increasing the possibility of interference by off-target amplification or primer dimer formation.

By comparison, sequence-specific probes provide greater specificity, while also allowing multiple gene targets to be analysed in the same reaction if different fluorescent reporters are attached to each probe. Such probes (e.g. TaqMan) consist of a short nucleotide sequence with a reporter and quencher attached to either end. The quencher prevents the emission of a fluorescent signal while the probe remains unbound, but when the probe is bound to its target gene during DNA amplification it is cleaved by the Taq polymerase to free the reporter and emit a detectable fluorescent signal.

Quantitative PCR data are plotted as cycle number *vs* fluorescence on a log scale (*Figure 2*) and are most commonly expressed as a threshold cycle (Ct; also referred to as Cq), which is a defined point at which the fluorescent signal rises significantly above background. The threshold is reached sooner if there is more DNA template to amplify, so the Ct is inversely correlated to gene abundance. Formal quantification by quantitative PCR may either be relative or absolute. Absolute quantification uses a standard curve derived from samples of a known DNA concentration; the concentration of DNA in an experimental sample is calculated by comparing its Ct to the standard curve. Less time-consuming and more common is relative

Figure 1. Quantitative polymerase chain reaction (PCR) couples DNA amplification to a fluorescent signal. a. 1. SYBR green emits minimal fluorescence when free in solution. 2. SYBR green binds to double-stranded DNA as it is amplified by PCR. 3. Once bound the dye emits a strong fluorescent signal which is detected by the quantitative PCR machine. b. 1. When sequence-specific probes are unbound the quencher (Q) prevents the emission of fluorescence by the reporter (R). 2. The probe binds to a specific nucleotide sequence as DNA is amplified by quantitative PCR. 3. The polymerase enzyme cleaves the probe as it amplifies DNA, freeing the reporter from the quencher and allowing it to emit a fluorescent signal which is detected by the quantitative PCR machine.



Mr James RC Miller *is MBPhD Student and* **Dr Ralph Andre** *is Senior Post-doctoral Research Fellow in the Department of Neurodegenerative Disease, UCL Institute of Neurology, London WC1N 3BG*

Correspondence to: Mr JRC Miller (j.miller@prion.ucl.ac.uk)

quantification, whereby the Ct value of a sample is expressed as a fold change compared to the Ct value of another sample.

The threshold cycle occurs during the exponential or log-linear phase of the reaction, where the amount of target DNA doubles with each cycle. This means that expression differences between two samples may be calculated as 2^x, where x is the difference between their Ct values. However, Ct values by themselves are often insufficient to compare gene expression, as they do not take into account additional factors such as differences in initial DNA concentration. This means it is necessary to normalize Ct values to those of reference genes which are expressed consistently between samples. This is known as the comparative Ct or delta delta Ct method, and can be summarized by the formula $2^{\text{delta Ct (sample 1 – sample 2)}}$, where delta Ct is the expression of the gene of interest normalized to that of the reference gene(s).

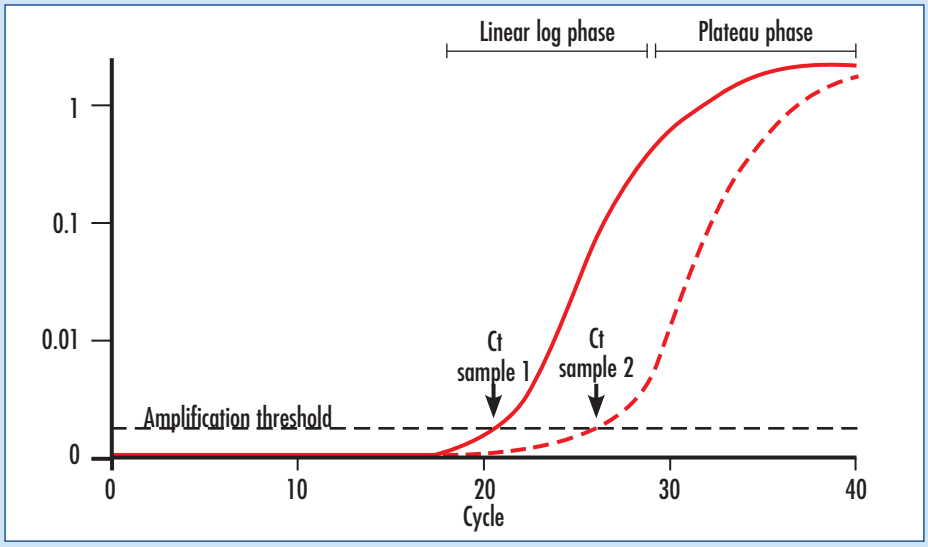


Figure 2. Example of a quantitative polymerase chain reaction amplification plot. The fluorescent signal is initially undetectable, but rises above background as DNA is synthesized – the threshold cycle (Ct) of a sample is the cycle number at which the fluorescent signal crosses the amplification threshold. In this example, sample 1 has a higher DNA concentration than sample 2 so it reaches the amplification threshold sooner. DNA is initially amplified exponentially, but eventually reaches a plateau phase in which the amount of DNA no longer doubles with each cycle as reagents are used and the polymerase becomes saturated.

Uses and applications

Quantitative PCR has a wide range of uses, the most common of which is the analysis of gene expression. This allows the investigation of how specific genes respond to differentiation, pharmacological agents and numerous other factors. Mostly the technique is limited to simultaneous quantification of small numbers of sequences, although dedicated quantitative PCR arrays are commercially available to quantify up to ~100 genes associated with a particular disease or biological pathway.

Quantitative PCR is also often used to validate specific sequences of interest identified by larger-scale exploratory analyses using microarrays or RNA sequencing. It can also be used to quantify mRNA knockdown following RNA interference experiments and to determine specific genotypes (e.g. for single nucleotide polymorphisms or the zygosity of a transgenic animal).

Clinical uses include the diagnosis of infectious diseases and cancer, the identification of genetic abnormalities, and the quantification of infectious agents in both patients and potential sources of disease (e.g. drinking water). Common examples range from the quantification and genotyping of the hepatitis B virus to the detection of bacterial resistance genes, for example the mecA gene which is associated with meticillin-resistant *Staphylococcus aureus*.

Strengths and limitations

Quantitative PCR provides a highly sensitive and specific means of quantifying a nucleic acid sequence. The assay has a wide dynamic range, does not require any post-PCR steps (e.g. gel electrophoresis) and is quick to perform, with a typical experiment capable of being set up and run in under 3 hours. This makes it an extremely valuable and much used research tool, but it has its limitations.

For example, quantitative PCR analysis relies on the assumption that DNA amplification occurs with 100% efficiency, i.e. that the amount of DNA doubles with each cycle. If the reaction takes place at less than 100% efficiency it can lead to considerable inaccuracies when the data are analysed, as increasing DNA concentrations will not be reflected in a proportionate decrease in Ct. Amplification efficiency should therefore be assessed using a dilution series of DNA. This may be calculated using the following formulae: E = 10^(–1/gradient); where gradient is that of the standard curve generated for reaction optimisation, and % Efficiency = (E-1) x 100%.

The difficulties involved in absolute quantification also mean that the vast majority of quantitative PCR experiments deal in relative differences and do not provide any exact information on DNA concentrations.

In biological terms, it is important to remember that gene expression is highly sensitive to external stimuli, and that experimental conditions must be carefully controlled to avoid confounding influences or external contamination. Moreover, there is no guarantee that the relative levels of mRNA measured will be translated into similar relative levels of functional protein, with numerous studies having demonstrated that mRNA and protein levels do not always correlate well (Gygi et al, 1999; Schmidt et al, 2007); gene expression data should be validated by protein analyses if the experimental hypothesis relies on protein expression differences.

Top tips and troubleshooting

Primer design

Primer design and validation is critical for quantitative PCR. Numerous software applications are available for designing primers, but there are a number of basic principles that should be followed (*Figure 3*).

While primer design is largely a theoretical exercise, the optimal annealing temperature can only be determined experimentally using melting curve analysis (*Figure 4*). This is a dissociation analysis done at the end of a quantitative PCR experiment to assess amplification specificity. A single

peak indicates the presence of a specific product, whereas multiple peaks indicate additional off-target amplification or primer dimer formation. During optimization a range of temperatures around the theoretical melting temperature of the primer set should be tested – the optimal annealing temperature will generate the lowest Ct value while avoiding any non-specific amplification as observed by the melting curve analysis. In practice, however, these steps are rarely required as pre-optimized primer sets are commercially available for vast numbers of gene targets.

Experimental design

Analysis of gene transcripts requires the synthesis of cDNA from mRNA. Such reverse transcription can either be done at the same time as quantitative PCR (one-step quantitative PCR) or separately before quantitative PCR (two-step quantitative PCR), but should be done with 100% efficiency to ensure a 1:1 ratio of mRNA: cDNA to accurately preserve expression differences. The final choice as to whether to use one-step or two-step quantitative PCR will depend on the experiment in question.

Quantitative PCR probe and primer concentrations are largely dependent on the DNA polymerase (check the manufacturer's instructions), although optimization is sometimes necessary. This is done empirically using a DNA dilution series to assess reaction efficiency – inadequate primer concentrations will be rate-limiting and result in reduced efficiency, while excessive primer concentrations will also inhibit the reaction as a result of increased off-target amplification and primer dimer formation. A reaction efficiency of at least 95% indicates that primer concentration is optimal. The remainder of the reaction components are usually supplied as a master mix with the DNA polymerase enzyme.

A typical quantitative PCR reaction consists of 40 cycles, each containing three steps. The DNA is first denatured at high temperature (usually 95°C) to separate it into single stands, before being cooled to allow primer annealing. The extension phase is then carried out at 70–72°C to allow optimal polymerase activity and is the point at which fluorescence is measured. If the amplicon is short it is common practice for the annealing and extension steps to be combined, as the polymerase is able to complete amplification during the ramp to the next denaturation step. The optimal length of each step will depend on

Figure 3. Features of effective primer and probe design.

Oligonucleotide primers are designed to amplify short segments of a target nucleic acid sequence in a polymerase chain reaction (PCR). Potential amplicons should be tested to select ones that have the highest signal-to-noise ratio (i.e. low threshold cycle with cDNA and no amplification with a no template control or genomic DNA). Good quantitative PCR primer design will have many of the following features:

The primer pair must be specific to the target gene; BLAST searching primer sequences will help determine the potential for off-target annealing

Shorter amplicons (50–150 base pairs) work most efficiently for quantitative PCR, although larger ones up to 200 base pairs can work if necessary

The optimal length for a quantitative PCR primer is 20 bases; the minimum number of bases is 18

The melting temperature (Tm) of the primer should be between 58 and 60°C. The Tm of a probe should be 10°C greater than that of its corresponding primers. If multiplexing, adjust probe length so that both probes have the same Tm

Primers should ideally contain approximately equal proportions of G, C, A and T bases; as a minimum, the GC content of primers and probes should be between 30 and 80%

Nucleotide repeats should be avoided where possible, in particular consecutive G residues should be limited to fewer than four

The five nucleotides at the 3' end of the primer should contain no more than two G and/or C bases, and G/C residues should be avoided at the 3' terminus. A is preferable to T at the 3' terminal, because thymidine tends to misprime more readily than other bases. G should be avoided on the 5' end of a probe because it specifically quenches the dye

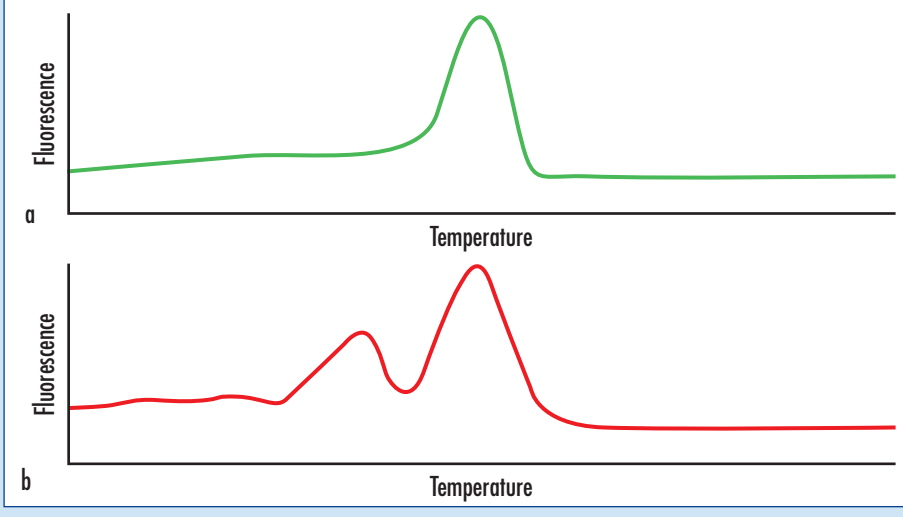
Primers should be chosen that do not form hairpins, do not form more than two bonds at the 3' terminus, and do not form stable homo- or hetero-dimers

If amplifying cDNA the amplicon should span one or more introns to avoid amplification of corresponding genomic DNA. If the gene you are studying does not have introns it may be necessary to run reverse transcription negative controls

Place the middle of the probe over the polymorphism for single nucleotide polymorphism analysis or over an exon–exon junction site. Primer to probe distance should not exceed 100 base pairs

Primer design can be aided by use of automated software that simplifies the process, allowing easy checks for confounding factors such as potential hairpins and primer dimer formation; examples include Primer Express, Oligo and Primer3

Figure 4. Example of a melting curve used to determine the specificity of a quantitative polymerase chain reaction experiment. a. The presence of a single peak indicates a single, specific product has been amplified. b. The presence of multiple peaks indicates that non-specific amplification has taken place, which may be caused by off-target primer annealing or primer dimer formation. In this case annealing temperature optimization or primer redesign is required to achieve assay specificity.



the enzyme in question, with some DNA polymerases being able to complete a run in as little as 40 minutes.

A quantitative PCR experiment should include sufficient biological and technical replicates, with at least three technical replicates being preferable so that outliers can be eliminated if one of the samples is not concordant. A 'no template control' should also be run to test for contamination with exogenous DNA, and will help to exclude the possibility of primer dimer formation. This is run in an identical manner to the experimental samples, but with water in place of the DNA sample; if the experiment is carried out correctly no amplification is seen.

Reference genes

Reference genes (also known as housekeeping genes) are used to normalize Ct values during relative quantification experiments. Selection of reference genes is a crucial part of any quantitative PCR experiment, and there are a number of criteria that must be fulfilled. The expression of a reference gene must be equal between samples, and its expression must be unaffected by whichever factor or disease process is being investigated. The reference gene must also have the same amplification efficiency as the gene(s) of interest. Failure to fulfil these criteria will lead to inaccurate normalization and it is good practice to use more than one reference gene to allow for unexpected variations. Which reference genes are used will depend on the experiment, and a thorough literature search should be conducted before their selection. Some common reference genes include GAPDH (glyceraldehyde 3-phosphate dehydrogenase), β-actin and β-2-microglobulin.

Amplification issues

Failure of amplification may be the result of a number of issues, from poor primer design to a problem with reverse transcription. As such it may be necessary to investigate each stage in the experimental workflow, and quality control for RNA purity and integrity is vital to ensure that the initial sample is not defective in any way. Amplification in the no template control is indicative of either exogenous DNA contamination or primer dimer formation; the cause may be elucidated by conducting a melting curve analysis as outlined above.

Conclusions

Quantitative PCR is an extremely powerful technique with wide-ranging applications from the analysis of gene expression to the diagnosis of infectious disease. The sheer number of available fluorescent markers, master mixes and thermal cyclers means the researcher's experience will vary considerably from laboratory to laboratory, but the principles outlined above remain applicable. While there are a number of potential sources of error, if quantitative PCR is designed, optimized and implemented properly it is a sensitive, reproducible and reliable technique which is a valuable tool in the arsenal of any medical researcher. **BJHM**

The authors would like to thank Richard Newton for his help with the graphics.
Conflict of interest: none.

Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**(3): 1720–30

Glossary

Absolute quantification – the use of a standard curve to determine the exact concentration of DNA in a sample

Amplicon – the DNA sequence between two primers which is amplified during a quantitative polymerase chain reaction (PCR)

Amplification efficiency – the proportion of template DNA which is amplified in each cycle of the quantitative PCR

Amplification plot – the plot of cycle number *vs* log fluorescence generated during a quantitative PCR

Baseline – the early stages of the quantitative PCR (usually cycles 3–15) where changes in the fluorescent signal are not detectable. This is used to set the threshold value

Biological replicate – samples taken from independent subjects in the same subject group to control for normal biological variability

Ct (or Cq) – the cycle number at which the fluorescent signal exceeds the threshold

DNA-binding dye – a non-specific dye which emits a fluorescent signal when bound to double-stranded DNA (e.g. SYBR green)

Dynamic range – the range of DNA concentrations which will give an accurate Ct value

Fold change – the relative difference between two values

Melting curve – dissociation analysis carried out after SYBR green quantitative PCR to validate assay specificity

Normalization – the use of reference genes to control for differences in starting DNA concentration

No template control – a reaction containing all of the quantitative PCR components apart from the DNA sample. Amplification indicates contamination or primer dimer formation

Primer – an oligonucleotide which binds to a specific target sequence to begin the process of DNA replication

Primer dimer – an unwanted PCR by-product occurring as a result of homology between primers

Probe – a specific nucleotide sequence with a fluorescent reporter attached to the 5′ end and a quencher attached to the 3′ end

Quencher – a molecule which prevents a probe in close proximity emitting a fluorescent signal

Reference gene (housekeeping gene) – a gene which is expressed consistently across samples and is used for normalizing Ct values

Relative quantification – the analysis of gene expression in one sample compared to another

Reporter – a fluorescent dye used to quantify DNA amplification

Reverse transcription – the process by which cDNA is synthesized from an mRNA template

Standard curve – a curve generated from samples of a known DNA concentration which is used as a reference during absolute quantification

Technical replicate – repeated analysis of a single sample to control for experimental variability

Template – the DNA sample which is amplified during a quantitative PCR

Threshold – a fluorescent signal above baseline which is used to determine a sample's Ct value

Schmidt MW, Houseman A, Ivanov AR, Wolf DA (2007) Comparative proteomic and transcriptomic profiling of the fission yeast *Schizosaccharomyces pombe*. *Mol Syst Biol* **3**: 79 (doi: 10.1038/msb4100117)

Further reading

Bustin SA, Benes V, Garson JA et al (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**(4): 611–22 (doi: 10.1373/clinchem.2008.112797)

Huggett J, Dheda K, Bustin S, Zumla A (2005) Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun* **6**(4): 279–84 (doi: 10.1038/sj.gene.6364190)

Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M (2010) A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. *Methods* **50**(4): S1–5 (doi: 10.1016/j.ymeth.2010.01.005)

KEY POINTS

- Quantitative polymerase chain reaction is a technique which allows quantification of the amount of nucleic acid in a sample of interest; it can be done in either an absolute or relative manner.
- It is commonly used for the analysis of gene expression, with a wide range of other uses in both basic science and diagnostics.
- The assay relies on the linking of DNA amplification to a fluorescent signal that increases in intensity as sequence-specific DNA is amplified.
- The cycle number at which fluorescence rises significantly above background is known as the Ct and is inversely correlated with the starting concentration of DNA in a sample.
- Some common issues include poor primer design, incorrect annealing temperature and contamination with exogenous DNA. They can generally be fully resolved with optimization and careful laboratory technique.