The MIQE Guidelines Uncloaked

8

Greg Shipley

Abstract

The MIQE (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) guidelines have been presented to serve as a practical guide for authors when publishing experimental data based on real-time qPCR. Each item is presented in tabular form as a checklist within the MIQE manuscript. However, this format has left little room for explanation of precisely what is expected from the items listed and no information on how one might go about assimilating the information requested. This chapter presents an expanded explanation of the guideline items with commentary on how those requirements might be met prior to publication.

Introduction

Led by Dr Stephen Bustin, a number of us with many years of experience in performing qPCR experiments contributed to the content within the MIQE manuscript (Minimum Information for Publication of Quantitative Real-Time PCR Experiment) that outlines the minimal requirements for the publication of real-time qPCR data and provides a checklist for reviewers and authors of new manuscripts containing real-time qPCR data (Bustin *et al.*, 2009). The purpose of this chapter is to help flesh out these requirements and hopefully better explain the thought processes behind the items listed. A copy of the checklist is available at the end of this chapter. When possible, as much information as possible should be provided either within the manuscript or as supplementary data available on-line for those readers interested in more in-depth information about how the technique was performed.

The performance of a real-time quantitative PCR (qPCR) experiment is not inherently difficult. Having said that, qPCR is a quantitative technique and therefore requires a degree of rigor in how the experiment is performed. Up until this year, qPCR data has been accepted for publication based on the strengths of the technique alone without questioning how carefully the various steps required for the experiment were carried out. The aim of the MIQE paper and guidelines is to provide the reader with all the information required to either repeat the experiment or be able to judge whether the data are sound. I'm sure everyone has experienced the frustration of looking at the Methods section of a paper to find some detail of how a qPCR experiment was performed only to find that there was scant information present. There are examples of research papers being retracted following

publication that should never have seen the light of day, as pointed out in the MIQE manuscript (Huang *et al.*, 2005; Bohlenius *et al.*, 2007).

It is beyond the purview of this chapter to cover all possible contingencies that might arise during a real-time qPCR experiment. However, an attempt will be made to offer some specifics or suggestions when possible for the various requirements as set out in the MIQE manuscript. I have avoided mentioning specific products, for the most part, as most of the reagents on the market today work quite well. However, there are a few cases where there is only one option and those products have been referenced. Another consideration in this decision is the observation that while the folks in laboratory A will have great success with product X over Y, those in laboratory B down the hall will swear that product Y is much better than product X. This illustrates what is most likely the most undocumented part of any research project, the effect of experience and the quality of the personnel involved on the outcome of the experiment. Since this factor can not be quantified in any reliable way, we have to depend upon the parameters outlined in the MIQE checklist as a measure of their success.. Unfortunately, there are many other examples of qPCR data in the literature that have been improperly or inadequately reported.

The following discussion will cover the main topics of the MIQE checklist in order. Hopefully, my commentary will help flesh out what we (the MIQE authors) had in mind when these items were put on the list.

Experimental design

The most important component of any experiment is in the planning. Like a chess match, looking several moves ahead prior to beginning can save a lot of time, energy and potential heartache down the road. One of the most important considerations is in determining how many members will be required in each experimental group to achieve statistical significance compared to one or more control groups. This is determined by performing a power analysis based on preliminary experimental data from a smaller subset of samples. It is important to define what constitutes an experimental and control group, the number (N) in each group and a clear representation of the data from each group.

As a Core Laboratory director, I insist that publications in which the data presented is generated within our facility mention our involvement in the acknowledgements or methods section of the paper. The 'coin of the realm' for Core Labs is providing data for investigators that lead to publications and grants.

Sample preparation

The term 'sample' in any experiment can have a large range of possible definitions. Therefore, it is critical that the author define what a sample is for each of their experiments. This will include a brief discussion of the origin of the sample (e.g. tissue culture, animal tissue, FFPE material, laser capture, blood, faeces, plant, microbe); how it was obtained (e.g. cell lysis, biopsy); how it was handled (e.g. flash frozen, RNAlater, homogenized immediately, fixation); and how it was stored prior to analysis (e.g. temperature and time prior to nucleic acid isolation). The amount of sample collected can also be critical when evaluating the RNA or DNA isolation procedure and should be reported. The collection of samples is arguably the most important phase of the experiment. If the nucleic acids are not stabilized sufficiently at this step, there is nothing you can do later on to rectify the situation.

It has been my observation over the years that it is important for investigators who have little experience with the collection, isolation and purification of nucleic acids to practice all phases of this procedure prior to initiating what may be a very costly set of experiments and ensure that they are technically up to the challenge. This is also true for personnel who have a lot of experience but are obtaining their material from a new source. Never assume that past methods will work with the new material.

Extraction of nucleic acids

Once samples are collected, the nucleic acids therein stabilized and stored appropriately, the next important step is the isolation of RNA, DNA or both. The method used is important and will be vary according to sample type. There is no 'one size fits all' for DNA or RNA isolation. Although the isolation of RNA or DNA is done primarily by hand, if one of the instruments available for automated sample preparation was utilized, this should be documented in the Methods. Usually a kit is used for nucleic acid isolation and if so the name, catalogue number and manufacturer of the kit should be mentioned. However, it is possible that homemade or additional reagents were employed during the purification process and these need to be completely described. A normal part of the isolation procedure will be the treatment of RNA with DNase I to remove a potential contamination source in the real-time RT-qPCR. Unlike RNase, *E. coli* DNase I requires Ca^{2+} for enzymatic activity and is stimulated by Mg^{2+}. Further, divalent cations will facilitate the cleavage of RNA with prolonged treatment at high temperature, even with no enzyme present. There are a number of methods and protocols for this step and the one used should be given. Treatment of purified DNA with RNase is common and should also be documented.

Once nucleic acid has been purified, its concentration, purity and quality are three critical parameters that must be measured and reported. The easiest way to measure the concentration is spectrophotometrically with an A_{260} measurement. An A_{260} value of 1 at a 1 cm path length equivalent will be 50 μg/ml for DNA and 40 μg/ml for RNA. However, this assumes that the nucleic acid is 100% pure, which is hard to achieve in practice. For that reason, an A_{260} measurement cannot be used to normalize real-time qPCR data for loading variations from sample to sample. For that purpose, a more specific fluorescence dye-binding assay, such as picogreen for DNA and ribogreen for RNA, is required to determine nucleic acid concentration. Since foreign proteins and other contaminants can cause PCR inhibition, it is critical that you report the purity of your nucleic acid preparations. The easiest way to accomplish this is by measuring ratios of light absorbance at different wavelengths. A ratio of A_{260}/A_{280} provides a measure of how much protein is in your RNA or DNA preparation. An often-cited minimum acceptable ratio is 1.8 up to the maximum of 2.2 for RNA or 1.8 to 2.0 for DNA but depending on the sample source, you may find that the minimum acceptable ratio is lower, although you must be able to show that these samples are working as well as cleaner samples. Another measure of cleanliness is the A_{260}/A_{230} ratio which will pick up salts and other contaminants. A ratio of 1.7 or higher is a good rule of thumb. For publication, a statement on the general range of these ratios for all samples used in the publication would suffice.

Of course, none of the above will tell you anything about potential RT or PCR inhibition but there are a number of ways to determine whether inhibition is a problem for your samples. One is to make 10-fold dilutions of the RNA or DNA and see if the samples

dilute properly (roughly a 3 PCR cycle decrement for each 10-fold dilution). You can spike in an *in vitro* transcribed RNA or a DNA oligo spanning the PCR amplicon for a gene or transcript not found in the sample and compare the Cq values when run alone or within a number of samples. For a more thorough analysis on PCR inhibitors see Chapter 2.

The last criterion is the most important, sample integrity. I have seen many samples that were abundant and clean of contamination but were also totally degraded and thus useless. The common method of determining RNA quality is by running the samples in a chip-based micro-electrophoresis chamber such as the Agilent 2100 Bioanalyzer or the Bio-Rad Experion instruments. Sample quality is based on an algorithm that takes the amount of the sample present in the 18S and 28S rRNA peaks as well as other factors to obtain a number that expresses RNA quality for that sample. It is also possible to use real-time qPCR assays for a number of transcripts made to both the 5′- and 3′-ends of each transcript. An ideal result would be if the amount found for the quantity of transcript at the 5′-end were identical to the amount recorded for the 3′-end of the same transcript. Thus, a ratio of 1 for the two values would be ideal but is not always found. However, the samples should have similar ratios within an experiment for each transcript (Nolan *et al.*, 2006a). DNA is much more stable than RNA during isolation so cannot be run on a microchip assay for analysis as total genomic DNA. But looking at DNA integrity on a low percentage agarose gel would be sufficient. No matter how you determine the integrity of your sample, the method used, along with the size range measured must be stated for all samples in the study (Nolan *et al.*, 2006b). A qualitative method that can give some indication of sample integrity is to observe the shape of the amplification curves following real-time qPCR. The amplification curves should all be parallel even when shifted to higher cycle numbers. If some samples have amplification curves that are less acute (upright) than the others or compared to a control template, PCR inhibition should be investigated using one of the methods described.

The reverse transcription step

The next most important step in an RT-qPCR experiment, after sample quality, is the reverse transcription of RNA into cDNA as most of the variability in Cq values will come from this step. Once the cDNA has been made, making replicates for the PCR becomes primarily a pipetting exercise, which is a common source of variability, but replicating the PCR alone does not constitute a true replicate of the RT-PCR. Thus, it is better to have multiple RT reactions, a triplicate RT per sample for example, with a single PCR from each RT reaction than to make a single cDNA reaction with a triplicate PCR.

There are many vendors that sell reverse transcriptase mastermix with two major kinds of reverse transcriptase, AMV and MMLV, and multiple ways to prime a cDNA reaction (assay-specific, random primers, oligo-dT and a mix of both). Thus, the combinatorial possibilities are rather large making it critical that the investigator describe in detail how the reverse transcriptase reaction was performed. The requirements are easily described but I will elaborate for each one.

1 The amount of RNA in micro- or nanograms and the volume used for the RT reaction should be given.
2 The gene-specific primer (GSP) is usually the reverse primer of the PCR assay but can be another primer more 3′ to the PCR amplicon. If the latter, the Tm is usually low

enough so it can not participate in the PCR but can anneal at the temperature used for cDNA synthesis. The annealing position within the sequence (5′-base position of the primer) along with the concentration of the primer must be stated.

3 The type of reverse transcriptase used, along with its concentration in units (or volume if units are not known) must be stated. I personally think that every investigator should know what kind of reverse transcriptase (AMV or MMLV) they are using as each has its own advantages.

4 Temperature and time used for the cDNA reaction simple but essential pieces of information that must be provided. The instrumentation used, whether a thermocycler, heater block and/or water bath, should also be stated..

5 When using a commercial mastermix the name of the mastermix, catalogue number and manufacturer should be stated.

6 The most important control for a successful RT reaction is the -RTase control. It is often questioned whether the Cq value of the –RT control has to be 40 cycles or no signal. The simple answer is the –RT control does not need to be negative. However, the Cq value for this control should be well shifted from the +RT reaction values. This is one of many examples of a situation where each investigator must decide on his or her own personal, hard and fast rule for data analysis and the rule should be reported. My personal rule here is a 6 Cq shift must be present between the mean Cq values of the +RT reactions and the -RT Cq value. Since a 3 Cq shift is roughly 10-fold, a 6 Cq shift equals a 100-fold difference between the RNA and any DNA contaminating signal. Thus, the RNA signal would contain at most 1% DNA contamination in the +RT Cq value. For me, that is acceptable. For you, it may not and you can make another rule. The key is to follow the rule no matter what and to state it in every publication. If data have to be thrown out, so be it.

7 Finally, you must state how the cDNA is stored prior to performing the qPCR. cDNA is fairly stable but it is important you treat it with care nonetheless and communicate how you have done so. At a minimum the storage temperature needs to be recorded and it is also useful to mention whether the sample was aliquoted prior to storage.

Assay template information

Whether you are working with transcripts or genes as primary targets, it is important to document each sample properly so the reader will know exactly to what you are referring. Common names for genes are not reliable and have evolved as more rigorous naming conventions have been employed to encompass all the many different genes, transcripts and variants found for many organisms. The following expands on the checklist requirements.

1 The gene symbol can be found at the NCBI, Ensembl, UCSC and many other databases available over the web. For example, if the 0estrogen receptor alpha was one of your gene targets, the gene symbol is ESR1.

2 Similarly, these sites will also provide an accession number which is a unique identifier for each gene (NC_xxxxxx), transcript (NM_xxxxxxx) or protein (NP_xxxxxx) sequence. The appropriate accession number should be provided in a table with other important components about the target (see Table 8.1).

Table 8.1 MIQE checklist for authors, reviewers, and editors

Item to check	Importance	Item to check	Importance
Experimental design		qPCR oligonucleotides	
Definition of experimental and control groups	E	Primer sequences	E
Number within each group	E	RTPrimerDB identification number	D
Assay carried out by the core or investigator's laboratory?	D	Probe sequences	D[d]
Acknowledgment of authors' contributions	D	Location and identity of any modifications	E
Sample		Manufacturer of oligonucleotides	D
Description	E	Purification method	D
Volume/mass of sample processed	D	qPCR protocol	
Microdissection or macrodissection	E	Complete reaction conditions	E
Processing procedure	E	Reaction volume and amount of cDNA/DNA	E
If frozen, how and how quickly?	E	Primer, (probe), Mg^{2+}, and dNTP concentrations	E
If fixed, with what and how quickly?	E	Polymerase identity and concentration	E
Sample storage conditions and duration (especially for FFPE[b] samples)	E	Buffer/kit identity and manufacturer	E
Nucleic acid extraction		Exact chemical composition of the buffer	D
Procedure and/or instrumentation	E	Additives (SYBR Green I, DMSO, and so forth)	E
Name of kit and details of any modifications	E	Manufacturer of plates/tubes and catalog number	D
Source of additional reagents used	D	Complete thermocycling parameters	E
Details of DNase or RNase treatment	E	Reaction setup (manual/robotic)	D
Contamination assessment (DNA or RNA)	E	Manufacturer of qPCR instrument	E
Nucleic acid quantification	E	qPCR validation	
Instrument and method	E	Evidence of optimization (from gradients)	D
Purity (A_{260}/A_{280})	D	Specificity (gel, sequence, melt, or digest)	E
Yield	D	For SYBR Green I, C_q of the NTC	E
RNA integrity: method/instrument	E	Calibration curves with slope and *y* intercept	E
RIN/RQI or C_q of 3′ and 5′ transcripts	E	PCR efficiency calculated from slope	E
Electrophoresis traces	D	CIs for PCR efficiency or SE	D
Inhibition testing (C_q dilutions, spike, or other)	E	r^2 of calibration curve	E
Reverse transcription		Linear dynamic range	E
Complete reaction conditions	E	C_q variation at LOD	E
Amount of RNA and reaction volume	E	CIs throughout range	D
Priming oligonucleotide (if using GSP) and concentration	E	Evidence for LOD	E
Reverse transcriptase and concentration	E	If multiplex, efficiency and LOD of each assay	E
Temperature and time	E	Data analysis	
Manufacturer of reagents and catalogue numbers	D	qPCR analysis program (source, version)	E
C_qs with and without reverse transcription	D[c]	Method of C_q determination	E
Storage conditions of cDNA	D	Outlier identification and disposition	E
qPCR target information		Results for NTCs	E
Gene symbol	E	Justification of number and choice of reference genes	E
Sequence accession number	E	Description of normalization method	E
Location of amplicon	D	Number and concordance of biological replicates	D
Amplicon length	E	Number and stage (reverse transcription or qPCR) of technical replicates	E
In silico specificity screen (BLAST, and so on)	E	Repeatability (intraassay variation)	E
Pseudogenes, retropseudogenes, or other homologs?	D	Reproducibility (interassay variation, CV)	D
Sequence alignment	D	Power analysis	D
Secondary structure analysis of amplicon	D	Statistical methods for results significance	E
Location of each primer by exon or intron (if applicable)	E	Software (source, version)	E
What splice variants are targeted?	E	C_q or raw data submission with RDML	D

[a] All essential information (E) must be submitted with the manuscript. Desirable information (D) should be submitted if available. If primers are from RTPrimerDB, information on qPCR target, oligonucleotides, protocols, and validation is available from that source.
[b] FFPE, formalin-fixed, paraffin-embedded; RIN, RNA integrity number; RQI, RNA quality indicator; GSP, gene-specific priming; dNTP, deoxynucleoside triphosphate.
[c] Assessing the absence of DNA with a no–reverse transcription assay is essential when first extracting RNA. Once the sample has been validated as DNA free, inclusion of a no–reverse transcription control is desirable but no longer essential.
[d] Disclosure of the probe sequence is highly desirable and strongly encouraged; however, because not all vendors of commercial predesigned assays provide this information, it cannot be an essential requirement. Use of such assays is discouraged.

3 The chromosomal location of the gene may or may not be important for the publication, but if it is, it should be provided, perhaps along with information on neighbouring genes or regulatory sites.
4 The PCR amplicon length (total span of the assay in bases) is useful information to provide. For example, if you are using FFPE samples and your SYBR Green I assays are all over 300 bases in length, it would not be a surprise that your ability to detect many transcripts was not very good.
5 If you run your PCR amplicon sequence on a BLAST search against the entire genome or transcriptome of the species in question, how many non-specific hits come up? Is there a reasonable expectation that both primers (and probe, if present) will allow you to detect the target sequence in a total RNA or gDNA population with high specificity?
6 Just because your assay crosses an exon/exon junction does not mean that you can not obtain amplification from contaminating DNA a priori. Pseudogenes are spliced copies of transcripts found in alternative places within the genome. Since they are occasionally, transcribed they often have mutational differences compared to the true gene. Known pseudogenes will show up on a BLAST search but not all of them are known. For this reason, it is best to keep DNA contamination to a minimum in RNA samples.
7 Prior to assay development, I always run an m-fold analysis to look for long stem structures than can prevent efficient binding of primers (or probes) to the target sequence. I find it best to avoid regions of high secondary structure during real-time qPCR assay development.
8 The location of exon boundaries within a transcript sequence is easily found these days for human, mouse and will be available for other species as progress is made on the analysis of each genome. Not all transcripts are spliced, of course. For easy identification, I find it is best to number each primer with the 5′-nucleotide base position within the refseq sequence and for probes we do the same. I also indicate which strand the probe binds to (+ or –) so the reader will know if the sequence is the same or the reverse complement to the reference sequence.
9 If the assay has been designed to detect one or a subset of splice variants known for a transcript, that information should be stated. If the assay will detect all know splice variants, we call that a generic assay but again this should be stated. Even if this information does not seem important at the time the paper is submitted [Please update if possible] it may have great relevance at a later date.

Assay components

In manuscripts where real-time qPCR is one of the techniques used, the sequences of the primers (and probe) used are invariably given. However, more detailed information concerning the assay past SYBR Green or probe-based is often missing. For proper review, it is critical that there be complete disclosure of the assay components and the location of the assay within the transcript or gene sequence. Further, comments on whether splice variants, allelic isoforms, SNPs, or other variations within the sequence are or are not detected with the assay should be disclosed if these issues are salient to the data presented within the manuscript. Assay component information is best presented in tabular form as

part of the Methods section. Table 8.1 shows the salient column headings and two example assays have been entered for reference.

The manufacturer of the primers/probe and oligo DNA standards, if appropriate, should always be given within the Methods section. Less important is the purification method used for the primers as the standard and free desalting for SYBR and probe-based assay primers is usually sufficient. There are instances, however, where HPLC purified primers are desirable. One example would be for primers used for high-resolution melt (HRM) analysis and this should be mentioned within the Methods section, although the jury is still out as to whether HPLC purified primers are in fact necessary for HRM. There are multiple synthesis and purification options for dual-labelled fluorescent hybridization probes so the method of purification as well as the reporter and quencher moieties should be given along with the manufacturer information. Lastly, there are base modifications that can be added to probes and primers and these need to be detailed (base position(s) and modification, e.g. LNA, inverted bases, etc.) along with the manufacturer information.

qPCR protocol

Describing the reaction conditions for a real-time qPCR experiment would seem to be an obvious addition to any Methods section when qPCR was used in data collection but there are publications where scant information on this topic is provided. Without this critical information it is nearly impossible for a reviewer to determine whether the data presented are credible. This part of the checklist outlines the basic information that must be stated in describing your protocol.

- If a commercial PCR mastermix was used: the name, catalogue number and manufacturer, along with a description of any modifications made by the investigator.
- If a home-made mastermix was used: a complete description of every component, manufacturer and final concentration,.
- The volume, amount of substrate added per reaction, plate type and geometry (48-, 96-, 384-well plates, and tubes for block instruments, capillaries or tubes for rotary instruments), model and manufacturer of the instrument.
- The thermocycling conditions.
- Although most experiments are set up by hand pipetting; if liquid handling robots are used that should be stated along with manufacturer and model.

qPCR validation

This section may be one of the most important in the checklist and yet is most often ignored in any Methods section and is generally given short shrift by investigators anxious to 'get the data' that they forget the data is only as good as the assay used to collect it. As stated earlier, after the quality of nucleic acid preparation, there is nothing more important than the qPCR assay and its' correct implementation in collecting data. All the fancy data analysis in the world can not make up for a poor assay. A good assay is joy to use forever and a poor one is a recurring nightmare.

The first step in new assay development is to acquire the information outlined in section 6 above. Next, determine where within the sequence the assay can and cannot be positioned to satisfy the research question at hand. All of these steps are accomplished *in silico*. This

may necessitate aligning related sequences, splice variants, related species, etc. to make this determination. Once the target area(s) have been identified, the actual assay design can commence. Whether the assay will use a probe (hydrolysis, beacon, scorpion, etc.) or use SYBR Green I or one of the many other possible dyes, a software program should be used. A list of assay design software is presented at the end of this chapter.

Two critical factors to keep in mind are that the primers are the most critical part of the assay and potential template secondary structure that can interfere with primer/probe binding. If your software of choice does not include secondary structure analysis as part of the design process, the sequence around the final assay PCR amplicon should be checked by m-fold analysis. Primers or probes that will anneal to strong stem regions within the target sequence will not work well and should be avoided.

Once the assay components have arrived in the lab, a preliminary PCR or RT-PCR should be run on a real-time instrument to see how well the assay is working. More importantly, the PCR product can be used to construct a standard curve. Even if you do not plan to run standard curves for data analysis, it is essential that one be run for assay quality control. The data from this experiment is part of the checklist information. Depending on the Cq value recorded for this first run, dilute the PCR product 100- to 1000-fold in 100 ng/µl *E. coli* or yeast tRNA (nuclease-free, molecular biology grade) in nuclease-free water. A 7-log range in 10-fold decrements from the initial dilution is usually sufficient to cover a high range down to 1–10 copies of template. From this experiment, you will find the dilution at which the assay no longer functions in a linear fashion with the higher template concentrations. The last dilution where the value was still linear (falls on the standard curve) is the LOD or lowest limit of detection. You can make 2-fold dilutions around this value to get a more accurate value if required. The critical consideration is that no Cq value can be reported if it falls below the LOD of the assay. If your assay is not good enough, one solution is to go back to the assay design software and find more primers that can work with the existing ones and try them in all possible combinations. Moving the primer over by 1 base can make a world of difference in how it works with a certain primer partner. I order 4 primers by default and do this for every new assay design. Another solution is to try the primers you have at different concentrations, both symmetric and asymmetric, and with different $MgCl_2$ concentrations and annealing temperatures. Once the results from these experiments gives satisfactory results, you can calculate the PCR amplification and PCR efficiency from the slope of the standard curve (see formulas at the end of this chapter). My QC limits are that the assay must be linear down to 10 template copies and have a PCR efficiency of 93% or better. If your assay does not meet these requirements but works for your experimental requirements, it can still be used. But, you will have to keep the limitations in mind for future work.

Some folks put a lot of stock in the r^2 measurement from the standard curve. I agree it should be near 1 but it is not the end all that some seem to think it is. It merely tells you that the person running the experiment can pipette accurately, which is important, but it is not the assay diagnostic some seem to think it is. A more telling value is the y-intercept but to get an accurate value for this term you have to accurately calculate the number of copies in a given mass of standard. When using PCR products, this can be hard to determine using an A_{260} measurement which is not all that accurate and is a sum of all components that absorb light at 260 nm. Using a value of 2×10^7 copies or molecules for a Cq value of 13 will get

you close to the desired y-intercept value. The y-intercept is a value that sums up all the important aspects of the assay (standard curve) in one value. A y-intercept value of 37–38 is within the desired range. A statement should be made in the Methods or Results section that all Cq values of unknowns fall within the linear quantifiable range of all the assays used.

If you have designed a multiplex assay (two or more assays within the same reaction), you must present data for each component assay in the multiplex as described above. Instruments have different ways of limiting spill over from the emission of one reporter dye into a s reporter dye that emits in near proximity (e.g. FAM and VIC or HEX) and this method should be mentioned in the Methods section.

Data analysis

The Data Analysis section primarily deals with how the final post-run data set was analysed. However, data analysis can be divided into two parts, immediate post-run analysis and post-run project data analysis. Immediate post-run analysis is queried under 'Cq Method'. Most instrument software depends on the user performing an initial analysis based on a thresholding process. That is a determination of the proper baseline and threshold to use to get the final Cq values for the run.

All software programs will do this automatically but it has been my experience, and one found in an unpublished ABRF study by the nucleic acids research group, that the best data analysis for all of the instruments in use at that time (2005) was when the end user set these two variables manually verses using any combination of manual and automatic or automatic analysis mode. I would encourage you to look at the manual settings for these two variables and see if you can make your data just a little better by setting these two variables yourself. It's easiest to determine this when running a standard curve over 6 or 7 logs but the proper settings will have an effect on how the Cq values of similar replicates appear as well. Having control of the threshold can be important when you want to compare samples across plates, for example. You should set both of these variables to the same value for all plates being compared, but you will find that the instrument software will not set them exactly the same and instead try to optimize them for each plate run. Not all instrument software works the same so it is impossible to give a generic recommendation here. Having said that, it is important that the baseline setting not include cycles with positive fluorescent signal from any of the samples and that the threshold is set high enough to avoid spurious low level PCR signals but not set so high as to being in the linear, verses the geometric amplification region of the PCR curve. Recently, Roche has introduced the 2nd derivative method of determining Cq values for their LC480 real-time instrument. This method eliminates the need for the investigator to intervene in determining Cq values. The new Bio-Rad CFX instrument uses a similar algorithm in their software. I mention this issue here because I fear most investigators do not know enough about how their real-time qPCR software works to have investigated these settings for themselves. However, if the initial Cq values are not the best they can be, succeeding data analysis can not rescue the data set.

The remaining parts of this section deal with the aspects that first come to mind when you think of data analysis. Although most of us have used Excel and some free-ware for the analysis of real-time qPCR data, there are some excellent programs on the market now that can provide a number of different analyses and save time to boot. The best of these are now commercial products but there are some free ones available as well. A web site that

lists all the available choices is one managed by Michael Pfaffl's group in Freising, Germany (http://www.gene-quantification.de/download.html), and an in-depth discussion of data analysis is found in Chapter 7. Regardless of which one you use, it should be referenced in the manuscript. The commercial software is also useful in determining when data points are statistically legitimate outliers. If you are performing an Excel -based manual analysis, the method for determining outliers must be given.

I think it goes without saying that data normalization is one of the most difficult parts of any real-time qPCR experiment. When a new experimental protocol is used without prior information on possible transcripts that could be used for data normalization, an empirical approach should be taken to determine possible candidates. One of the goals of the MIQE paper was to eliminate the use of the term 'housekeeping gene' whose very definition was 'a gene that was constitutively transcribed and invariant in concentration regardless of circumstances'. We now know that there are no genes that fall into that category and that any transcript, including the main ribosomal genes 18S and 28S rRNA, can vary depending on the experimental conditions. Why do the data need to be normalized, you might ask? The simple answer is that absorbance estimations of nucleic acid concentration are not sufficiently accurate to be used for data normalization. Thus, when you think you have added the same amount of RNA by absorbance assay for 20 samples into an RT-qPCR experiment, in reality you will find that there will be about a 10-fold spread in the actual measured amounts for any 'control or normalizer' transcript, even if all the samples are from control biological replicates. This just emphasizes the variability in determining macromolecule concentrations using an absorbance assay. Thus, it is critical that something besides absorbance be used to normalize for the amount of sample going into each RT-qPCR or qPCR. Following the seminal publication by Vandesompele *et al.* (2002), it was clearly shown that normalization of RT-qPCR data was best done using multiple relatively invariant transcripts rather than a single one. In practice, any transcript can be used for data normalization if it is relatively invariant in all samples within that experiment. The converse of that has been seen as well where transcripts that one thought would be used for data normalization turned out to be used as experimental ones due to the changes in that transcript within the experiment. It is truly the case that you do not know until you know what will work for data normalization and it is critical that this determination be made early during the experimental development phase. There are programs that can help identify the best transcripts or genes to use for data normalization from empirical data. Software that is dedicated to data normalization is geNorm from the Vandesompele lab (http://medgen.ugent.be/~jvdesomp/genorm/) or NormFinder at the Nordic Centre of Excellence in Molecular Medicine (http://www.mdl.dk/publicationsnormfinder.htm). A number of vendors sell premade plates with multiple potential normalizers on them for rapidly screening and determining which ones to use for a new experimental condition.

It is also true that it is not always possible to rely on any transcript or gene for data normalization. Some examples I have encountered are actinomycin D-treated cells for transcript half-life experiments, cells undergoing apoptosis, and genes or transcripts that originate outside the context of a cell such as serum, plasma, saliva, urine, pus, faeces or other biological fluids. I am sure there are other good examples from microbiological or ecological studies as well. In these cases the best option is to measure the total nucleic acid in the sample using a fluorescence assay such as picogreen for dsDNA, oligreen for ssDNA

and ribogreen for RNA (Molecular Probes/Life Technologies, Carlsbad, CA, USA). The final data would be expressed as molecules or copies/ng total RNA or DNA. Copy number can be interpolated from a standard curve or relative quantities can be calculated from Cq values using the PCR amplification value for the assay in the formula (amplification)$^{-dCt}$, where an ideal real-time qPCR assay would have the formula, 2^{-dCt}.

Another item that is important is the difference between a technical replicate and a biological replicate. Technical replicates are repeated analyses of the same sample. That is multiple PCRs from the same cDNA pool or performing replicate RTs from a single sample and a single PCR from each RT reaction (preferred). In both cases the final values will be averaged and a standard deviation calculated (if triplicates or above are determined) but the value will still represent a statistical $N = 1$ for that data point. Technical replicates give you more accuracy in measuring the Cq value for a single transcript on a single sample. A biological replicate is a collection of different samples, all treated the same. So, if you have five control samples, you might run a triplicate technical replicate for each one but at the end of the day, you would still have an $N = 5$, one mean value for each sample which can then be averaged to give the final mean for all the samples in that experimental group. Biological replicates give you more confidence that the data point you end up with is as close as possible to the 'real' value for that experimental treatment within the total cellular or organismal population. Bottom line, it is important to report the number of technical and biological replicates used for data analysis. From this data set, it is important to report the intra-assay variability for each transcript or gene determination. This is best done by showing mean, standard deviation and %CV for each assay from a single plate or run determination. If multiple plates or runs are used to obtain the complete data set for each transcript or gene over the entire set of samples, inter-assay variability should be presented as well.

Finally, the method used to determine statistical significance for your data set needs to be reported. If you performed a power analysis to determine how many members would be required to obtain significant differences in each experimental group, that information should be provided as well along with the software or formulas used for these calculations.

Final comments

The overarching goal of the MIQE requirements is the full disclosure of how a real-time qPCR experiment was performed so that the reader can either repeat the experiment(s), if desired, or at least know exactly how they were performed. Once assays are completely disclosed in the literature and the procedures outlined for an investigator, it should be possible to refer back to the seminal paper in future manuscripts if, and only if, no changes in the assays and procedures have been made in succeeding experiments. If some modification has been made, a description of the modified portion of the experimental protocol should be reported.

It has become common practice to pare the Methods section down to collections of partial and in some cases bare bones information that is insufficient if one wishes to repeat the experiment or just be able to determine if it was done properly. This was brought on by pressures to keep the number of pages to a minimum due to the main factor that rules most things, cost. With the advent of the web, we can now put the missing detailed information that has heretofore been lacking in supplemental information sections that are readily

accessible by the reader and yet still keeping the economic realities of paper publication in check.

Satisfying these requirements may seem to be rather daunting but in fact they are quite manageable and more importantly just plain good science. Hopefully the information in this chapter has answered questions you may have had about the publication of real-time qPCR data. As with any experimental protocol, there are many wrong but few right ways to perform the technique. With the goal of having solid real-time qPCR data you want to share with the rest of the research community, if you follow the MIQE guidelines you will succeed.

Useful websites

NCBI (http://www.ncbi.nlm.nih.gov/)
m-fold (http://mfold.bioinfo.rpi.edu/)
ensembl (http://www.ensembl.org/index.html)
UCSC genome (http://genome.ucsc.edu/)

Primer design programs

Product	Company/source	Web site
Commercial		
Beacon Designer	Premier Biosoft	www.premierbiosoft.com/
AlleleID	Premier Biosoft	www.premierbiosoft.com/
3Visual OMP™	DNAsoftware	http://dnasoftware.com/
Primer Express	Applied Biosystems	www.appliedbiosystems.com
Free		
RealTimeDesign	Biosearch Technology	http://www.biosearchtech.com/realtimedesign.aspx
Primer3	MIT	http://frodo.wi.mit.edu/primer3/
PCR Design Tool	IDT	www.idtdna.com/scitools/applications/realtimePCR

Real-time qPCR analysis programs

Product	Source	Web Site
GenEx	MultiD	http://www.multid.se/genex.html
qBasePlus	Biogazelle	http://www.biogazelle.com/products/qbaseplus
REST-2008	Gene-Quantification	http://www.gene-quantification.de/download.html

Useful real-time qPCR formulas

Assay amplification = $10^{-(1/\text{slope})}$

The ideal value here is 2, for one template converted to two within one cycle.

Assay PCR efficiency = (assay amplification –1) × 100

The ideal here is 100%: (2 – 1) × 100 = 100%.

Relative quantity for a sample = $(\text{amplification})^{-\text{dCt}}$

Relative quantity for a sample in relation to another sample = $(\text{amplification})^{-\text{ddCt}}$

Molecules in 2 pg of ssDNA template = $\{2 \times 10^{-12}/[\text{DNA length in bases} \times 330 [\text{average MW of dNTP})]\} \times 6.023\text{E}23$

In 2 pg of a 78-base ssDNA template there would be 4.68×10^7 copies

Molecules in 2 pg of dsDNA template = $[2 \times 10^{-12}/(\text{DNA length in bases} \times 660)] \times 6.023\text{E}23$

Molecules in 2 pg of an RNA template = $[2 \times 10^{-12}/(\text{RNA length in bases} \times 340)] \times 6.023\text{E}23$

References

Bohlenius, H., Eriksson, S., Parcy, F., and Nilsson, O. (2007). Retraction. Science *316*, 367.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., and Wittwer, C.T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem *55*, 611–622.

Huang, T., Bohlenius, H., Eriksson, S., Parcy, F., and Nilsson, O. (2005). The mRNA of the Arabidopsis gene FT moves from leaf to shoot apex and induces flowering. Science *309*, 1694–1696.

Nolan *et al.*, 2006a, Nat Protoc, *1*, 1559–82.[Please supply complete reference]

Nolan, T., Hands, R.E., Ogunkolade, W., and Bustin, S.A. (2006b). SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. Anal. Biochem. *351*, 308–310.

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.