Biomolecular Detection and Quantification 14 (2017) 19–28

Contents lists available at ScienceDirect

Biomolecular Detection and Quantification

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

Review Article

qPCR primer design revisited

Stephen Bustin[a,*], Jim Huggett[b,c]

[a] Faculty of Medical Science, Anglia Ruskin University, Chelmsford CM1 1SQ, United Kingdom
[b] Molecular and Cell Biology Team, LGC, Queens Road, Teddington, Middlesex TW11 0LY, United Kingdom
[c] School of Biosciences & Medicine, Faculty of Health & Medical Science, University of Surrey, Guildford, GU2 7XH, United Kingdom

ARTICLE INFO

Handled by: Justin O'Grady

Keywords:
Real-time PCR
Assay design
MIQE
Oligonucleotides

ABSTRACT

Primers are arguably the single most critical components of any PCR assay, as their properties control the exquisite specificity and sensitivity that make this method uniquely powerful. Consequently, poor design combined with failure to optimise reaction conditions is likely to result in reduced technical precision and false positive or negative detection of amplification targets. Despite the framework provided by the MIQE guidelines and the accessibility of wide-ranging support from peer-reviewed publications, books and online sources as well as commercial companies, the design of many published assays continues to be less than optimal: primers often lack intended specificity, can form dimers, compete with template secondary structures at the primer binding sites or hybridise only within a narrow temperature range. We present an overview of the main steps in the primer design workflow, with data that illustrate some of the unexpected variability that often occurs when theory is translated into practice. We also strongly urge researchers to report as much information about their assays as possible in their publications.

1. Introduction

The peer-reviewed literature contains references to tens, if not hundreds of thousands of oligonucleotide primer sequences for use with the polymerase chain reaction (PCR) and hundreds more are available from primer databases or can be bought from commercial suppliers. Oligonucleotide synthesis is available at bargain prices, enzymes are becoming faster, more reliable and cheaper, there are task-specific master mixes (e.g. for multiplexing) and thermal cyclers are becoming more affordable and user-friendly. This makes it possible to generate huge amounts of data with comparatively little effort. Since these data find their way into over 15,000 qPCR-related publication every year, it is essential to try and ensure that publications report real results, rather than technical bias [1].

With so many ready-made assays available, one might wonder why anyone would want to go to the trouble of designing yet another assay. Especially as the perception is that designing one's own assay is a lot more complex and inconvenient than simply buying it from a commercial supplier, who in any case will have validated every one of their assays. That perception is wrong for two reasons: first, commercial primers or assay condition may not have been experimentally validated or optimised. Second, it cannot be presumed that a primer set will generate the same results under different experimental conditions since assay performance can vary depending on what extraction methods was used to purify the templates [2], what reagents were used for the PCR reaction [3–5] and what thermal cycler was used to run the assays [6,7]. Hence researchers can be sure of an assay's performance only by performing their own validation and optimisation experiments. Doing this before working with precious samples will save time, expense and help avoid failed runs or inconsistent experimental data. Given the significance of empirical validation, it is important that any publication include that essential information [8–11]. Several reports have been published recently that together scored thousands of peer-reviewed papers in a wide collection of journals ranging from low to high impact factors [12–17]. All concluded that the amount of critical information provided with papers reporting qPCR data is inadequate for the purpose of evaluating the validity of conclusions arising from those data, with many not reporting primer sequences, validation data or including wrong information.

The main concerns with regards to designing assays usually relate to researchers being unfamiliar with the primer design process or unsure about the key parameters most likely to generate optimal primers, lacking the appropriate design tools and apprehension that the design process will take too long. However, assay design is usually quite straightforward, suitable tools are freely available online and it takes less time to design a robust, sensitive and specific assay than to troubleshoot a poorly designed one.

We provide a concise overview of the main primer-related issues

* Corresponding author.
E-mail addresses: stephen.bustin@anglia.ac.uk (S. Bustin), jim.huggett@lgcgroup.com, j.huggett@surrey.ac.uk (J. Huggett).

https://doi.org/10.1016/j.bdq.2017.11.001
Received 24 March 2017; Received in revised form 8 November 2017; Accepted 12 November 2017

that confront anyone wanting to design a qPCR assay, consider the main criteria that have an impact on assay performance, dissect the individual steps of the assay design workflow and analyse the performance of some real-life assays.

2. The importance of primers

Appropriately validated primers are crucial in determining the specificity, sensitivity and robustness of a PCR reaction [18]. Whilst it nearly always possible to get a result with a PCR assay, this is not the same as getting a correct result, be that a present/absent call for the detection of a pathogen or mutation using an endpoint assay or an accurate quantification of RNA copy numbers using a real-time method. In reality, PCR is not as robust as many people believe and there is a need to consider the science underlying DNA folding and match versus mismatch hybridisation. Having said that, it is not always obvious why some primer combinations work, or indeed do not work well.

The critical variable for primer performance is its annealing temperature (T_a), rather than its melting temperature (T_m), as the T_a defines the temperature at which the maximum amount of primer is bound to its target. The optimal primer T_a must be established experimentally as primer design programs generally calculate T_ms and, in any case, many use wrong prediction parameters [19]. Furthermore, since optimal annealing temperatures vary with different buffers, results obtained with one master mix cannot necessarily be extrapolated to a second one. Even at the optimal T_a, non-specific amplification can occur, especially with "proofreading" enzymes, caused not just by primer dimers but by physical closeness of primer pairs at mismatched sites. Furthermore, reliance on BLAST searches alone does not guarantee primer specificity, since whilst the BLAST algorithm returns fast results it may miss thermodynamically important hybridisation events as it does not correctly score the gaps that generate duplex bulges [19]. Furthermore, the effects of mismatches on duplex stability are sequence context dependent and are not correctly called by sequence independent approximations [20].

3. Principal considerations for assay design

A good assay will not create primer dimers, be close to 100% efficient and exquisitely specific. Such an assay will also be robust, which means that if conditions are not quite optimal, for example if a sample contains traces of an inhibitor, or if the thermal cycler has uneven thermal profiles across its block, then the assay may still perform reliably and generate usable data. In contrast, a poor assay will be much more susceptible to variable conditions, and is virtually guaranteed to result in wasted time and considerable frustration on the part of the researcher. As a rule of thumb, if primers perform well over a broad temperature gradient, the assay tends to be robust, whereas if amplification is restricted to a narrow temperature optimum, it is not.

When designing assays in-house, the design process comprises a comprehensive workflow that demands careful consideration not just of the primers themselves but also of amplicon uniqueness, structure and location, with the aim of bringing about an optimal primer/amplicon combination for accurate quantification of nucleic acids. Attention to such detail makes it more likely that the assays will yield data that are sufficiently reliable and sensitive to generate consistent as well as biologically/clinically relevant results. Importantly, even when the primers have been designed by a colleague, tracked down from a peer-reviewed publication, acquired from a primer database or purchased from a commercial source, reliable qPCR demands a retrospective evaluation of most of the *in silico* criteria and assiduous validation of all of the wet lab parameters.

Achieving these objectives is not difficult, when following the workflow shown in Fig. 1 which involves four major steps: (i) target identification, (ii) definition of assay properties, (iii) characterisation of primers and (iv) assay optimisation. The first two steps are carried out by *in silico* analyses, the latter two by experimental investigation

4. Target identification

It is self-evident that an assay is useful only if the correct target has been identified and used for assay design. Hence the more that is known about the DNA or RNA of interest, the better. Accordingly, the first step involves accumulating as much information as possible from sequence databases. This can appear to be quite daunting, but will soon become second nature with a systematic, step-by-step approach. One problem with searching for sequences is that there are often numerous identical, closely-related or, more surprisingly, significantly different sequences listed under the same common name. An NCBI search for "*Aspergillus terreus* 18S" brings up 161 sequences of varying lengths, descriptions and accession numbers. A search for "*Aspergillus terreus* 28S" returns 142 sequences, some the same as the previous search, but again all with different accession numbers and lengths. There are two lessons here: (i) it is essential to have absolute clarity about the amplification target and (ii) it is crucial always to refer to the accession or individual transcript number of any sequence used for assay design, as this minimises the risk of confusion and makes life much simpler for reviewers and readers. Furthermore, many databases are not curated, so a given sequence name is solely based on what the individual who uploaded it thought it was. Consequently, if the original sequence was incorrect, for example due to a nonspecific PCR, but this was not known when it was uploaded, then a circular problem will arise that further propagates the error.

Database mining assumes familiarity with its nomenclature: e.g. NCBI sequences prefixed with NC_, NG_ are curated genomic sequences, NM_ is a curated mRNA sequence, whereas NT_ and NW_ are automated genomic and XM_ automated RNA sequences. Hence the information with regards to some sequences (NM_) is likely to be more reliable than that for others (XM_) and judicious choice of sequence information is advised. For example, if the aim is to amplify a cellular mRNA, it is important to ascertain whether there are transcript or splice variants or additional closely related paralogues and whether the assay should target all of those or be able to distinguish between them. One important consideration is to ensure that the assay does not inadvertently amplify pseudogenes. Designing PCR primers for miRNAs is somewhat more challenging, since a typical miRNA is only 22 bases long, which is about the same size as a conventional PCR primer. Genotyping assays, on the other hand, place an obvious restriction on the position of the amplicon as it must include the site of the polymorphism or mutation. This highlights an important point in that while designing an optimal assay is desirable the designer is ultimately at the mercy of the sequence in question and the 'best' assay may not be ideal. This may not preclude its use as long as the limitations of such a choice (such as possible reduced efficiency, sensitivity or precision) are understood and, crucially, reported in any publication.

Absolute certainty as to what is being targeted is of particular importance when utilising PCR as a diagnostic or forensic test. Whilst assays used in clinical applications are heavily regulated (although mistakes are still made), research diagnostic assays are not, so factors such as specificity must be considered when reaching conclusions. Human papillomavirus (HPV) is the primary aetiological factor that transforms cervical epithelia into cervical cancer and causes most anal and oropharyngeal as well as some vaginal, vulvar, and penile cancers. According to epidemiological case-control studies, 15 high-risk HPV types have been acknowledged, while three types have been designated as probable high-risk and 12 types have been classified as low-risk [21]. Depending on the purpose of an associated study it could be essential to distinguish between these subtypes and so appropriately designed assays will be paramount [22]. Similarly, designs targeting bacterial and fungal pathogens require careful consideration prior to carrying out any diagnostic experiments [23]. The recent interest in using RT-PCR to target RNA for the tissue profiling of human forensic samples [24]