



ProbeLibrary: A new method for faster design and execution of quantitative real-time PCR

Peter Mouritzen, Mikkel Noerholm, Peter S Nielsen, Nana Jacobsen, Christian Lomholt, Henrik M Pfundheller & Niels Tolstrup

ProbeLibrary, a fast, specific and flexible format for quantitative real-time PCR, is described. The ProbeLibrary concept is based on the fact that just 90 short probes provide transcriptome-wide coverage in most organisms. These short probes are highly specific and possess the high melting temperature (T_m) required for real-time PCR owing to the incorporation of locked nucleic acid (LNA). The probes are dual-labeled fluorogenic probes that are used in standard real-time PCR protocols, with standard instrumentation. ProbeLibrary probes are used in assays designed using free, web-based assay design software, termed ProbeFinder. In seconds, ProbeFinder designs highly specific intron-spanning assays for a target transcript. The system allows multiple design options, including designs for transcript variants and gene family members. Assay design success rate is 96%. This combination of extremely fast assay design and instant availability of the probes leads to faster implementation of real-time PCR assays.

Complete or near-complete genomic sequences have become available for an increasing number of different organisms. The availability of genomic information has generated a growing interest in understanding the dynamics of genetic regulation in basic biological processes, disease diagnostics, toxicology, drug screening and development. In studies of gene expression levels, technologies such as DNA microarrays and real-time quantitative RT-PCR (qPCR) have essentially replaced traditional approaches such as northern blot analysis and RNase protection assays. Because qPCR has unparalleled dynamic range and extremely high sensitivity, it has rapidly become the method of choice for validation of microarray data and gene knockdown experiments.

Challenges of real-time qPCR

Although qPCR is, in principle, a simple yet powerful technique, the typical researcher is still faced with several challenges. These include choosing between flexible or specific assay formats, designing the assay, waiting for the arrival of assay components, and actually performing the assay. The ProbeLibrary concept—90 qPCR probes in combination with assay design software—offers several possible solutions to these challenges.

Choosing between flexible or specific assay formats:

ProbeLibrary assays are both flexible and specific

Current approaches for carrying out qPCR fall into two broad categories: intercalating dye-based (predominantly SYBR Green) and

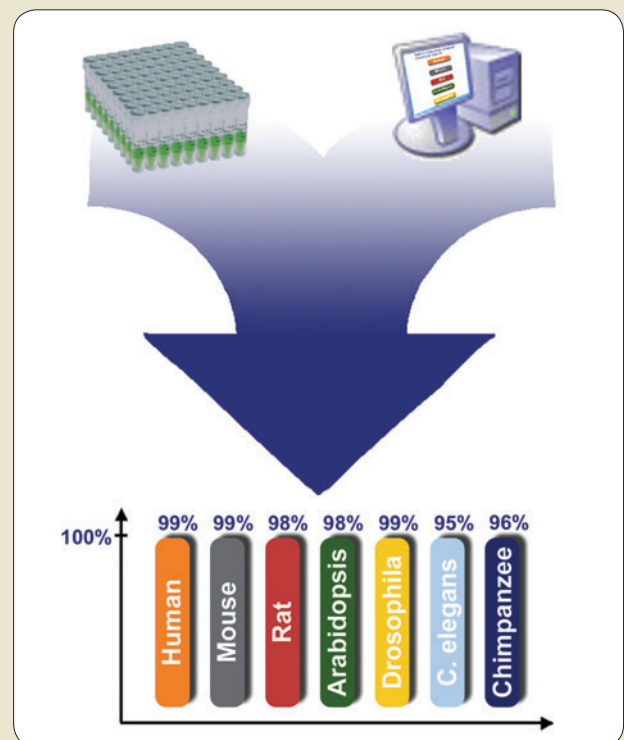


Figure 1 | The ProbeLibrary concept: 90 prevalidated real-time PCR probes and the freely available software ProbeFinder are used to design custom assays in seconds, enabling the quantitation of almost any transcript the day after the design of assays.

APPLICATION NOTES

probe-based (for example, hydrolysis probes). SYBR Green-based assays are considered very flexible as they only require the design of a primer pair, but they are generally less specific than probe-based assays because nonspecific amplification and the formation of primer dimers can result in erroneous estimates of expression levels. By contrast, probe-based assays ensure high specificity through the combination of gene-specific primers and the qPCR probe. But the probe-based assay format is less flexible, requiring the design of a primer pair and a specific functional probe to form an effective assay.

The ProbeLibrary exploits the fact that certain 8- to 9-mer sequence motifs occur with very high frequency in the transcriptome of many organisms and are spread throughout the entire transcriptome. Each of the 90 probes in a ProbeLibrary is a short (8- or 9-nt), prevalidated, dual-labeled probe that has been selected to target one of these frequently recurring sequences in the transcriptome of the relevant organism. For example, 99% of all transcripts registered in the human gene RefSeq database at the National Center for Biotechnology Information (NCBI) are tar-

geted by at least one of the 90 probes in the ProbeLibrary (Fig. 1). Furthermore, 90% of all transcripts in the Ensembl human gene database are targeted near an exon-exon boundary, enabling the design of intron-spanning assays for almost any intron-containing gene. Each transcript is targeted, on average, by 19 different probes. This allows the design of several assays per transcript, providing a means to cross-validate assays and increase confidence in results. Moreover, specific assays can often be designed to quantify different splice variants or individual gene transcripts in gene families. The availability of several assays per transcript also limits the amount of optimization required. In case an assay is encountered that performs inadequately, selecting an alternative assay is usually a faster way to achieve results than spending time on optimization.

The functionality of the very short 8- to 9-mer probes has been obtained by substituting individual nucleotides in the probes with the high-affinity DNA analog LNA¹. The use of LNA increases the T_m of the probes to facilitate standard real-time PCR cycling with annealing and extension at 60 °C. Moreover, the use of LNA increases the base-pairing specificity of the probes relative to DNA, and therefore the combined use of a gene-specific primer pair and the probe ensure high assay specificity. Each individual probe has been subjected to thorough testing and validation in real-time PCR to ensure optimal performance before inclusion in a ProbeLibrary.

Designing the assay: Assay design is fast and simple using the free, web-based ProbeFinder software

Designing a qPCR assay can be complicated, requiring identification of the correct transcript sequence, comparison with homologous gene sequences, determination of the exon-intron structure and identification of splice variants. After completion of these tasks, specific primers must be designed, which requires BLASTing primer sequences against genome and transcriptome databases to ensure uniqueness. The BLAST results must be deciphered, and for probe-based assays the task of designing a functional probe remains, which can also be a complex process depending on the probe type.

The free ProbeFinder assay design software is an integrated part of the ProbeLibrary concept. The user enters the transcript sequence of interest through a simple and user-friendly interface, and ProbeFinder generates an intron-spanning assay, consisting of PCR primers combined with a selected ProbeLibrary probe. The ProbeFinder software identifies the exon-exon boundaries within submitted transcripts and by default attempts to design intron-spanning amplicons for real-time PCR assays. The intron identification is performed in a number of ways: either by a look-up in the Ensembl database, by user-defined positions marked in the submitted sequence, or by an intron prediction algorithm based on a BLAST search of the submitted transcript against the relevant genome sequence. Having identified intron sites, ProbeFinder searches the submitted transcript for target sites for the 90 ProbeLibrary probes. Primer positions are designed to favor short (<150 bp) intron-spanning amplicons. Possible primer candidates are subjected to an *in silico* PCR algorithm that searches genome and transcriptome sequences to ensure a high degree of assay specificity. The assays that are most likely to show good performance are then identified and ranked, based on the *in silico* PCR-determined specificity, the amplicon length and whether assays are intron-spanning or not. The ProbeFinder 2.0 software accepts submission and parallel analysis of several transcripts (for example, different splice variants or gene family members), and assays may be identified that will differentiate between these (Fig. 2). Instead of the specific gene

The screenshot shows the ProbeFinder software interface. At the top, it says "EXIQON ProbeLibrary™". Below that, it states "This is ProbeFinder™ for the Human ProbeLibrary™". The main content area shows "Use ProbeLibrary™ probe(s): Human#01, Human#13 and Human#20". There are three transcript entries, each with a table of probe details and a graphical representation of the transcript with primer positions.

Input	Probe ID	Primer	Length	Position	Tm	%GC	Sequence
ENST00000211373.1	Human#01	Left	18	758 - 775	60	56	tcagcatgtgtctgga
		Right	20	813 - 832	59	50	ccactgtgagccagcatt
ENST00000211363.1	Human#13	Left	18	752 - 769	60	61	ccctccagctgtctct
		Right	21	801 - 821	60	48	aactgaggaatccagcatt
ENST00000229504.1	Human#20	Left	18	102 - 119	60	56	ttcagcaagggctgtac
		Right	18	187 - 204	59	44	tttccgcccctccaaaa

Figure 2 | Assay design page from the ProbeFinder software, showing the design of three assays that allow discrimination of three splice variants for the human C6orf21 (also known as G6f) transcript.

ID number, ambiguous gene names (for example, “MAPK” or “MAP kinase” for *MAPK*) can also be submitted. In this case, the software presents a list of GenBank entries, from which the user can select gene family members or splice variants for design of discriminating assays.

Waiting for arrival of assay components: ProbeLibrary eliminates delivery time for real-time probes

Delivery time affects the number of new assays that can be implemented in the laboratory per year and consequently has a large impact on the generation of results. Among the different components required to run a real-time PCR assay, probes tend to have the longest delivery times, because synthesis and purification is time-consuming for providers of these dual-labeled oligonucleotides. If the designed real-time PCR probes do not function properly, redesign is required, followed by waiting for synthesis, purification and delivery. Prevalidated assays also have relatively long delivery times, often of 2–3 weeks.

As the ProbeLibrary consists of only 90 probes, all the required probes for analysis of virtually any gene are easily stored in a standard laboratory freezer. Consequently, probe synthesis and delivery delays are eliminated. The PCR primers required for a particular assay can often be synthesized overnight and delivered the next day. This dramatically cuts down the implementation time for new assays and increases throughput.

Performing the assay: ProbeLibrary assays are fully compatible with existing real-time PCR instruments and protocols

Successful performance of an assay requires full compatibility with existing real-time PCR instruments and commonly used assay components and chemistries.

The ProbeLibrary probes are labeled with fluorescein (FAM) and are thus read by all standard real-time PCR cyclers. The assays are compatible with standard *Taq* polymerases and ready-to-use master mixes. The T_m of the probes has been adjusted by inclusion of LNA to ensure compatibility with standard real-time PCR cycling conditions and protocols.

Performance of ProbeLibrary assays

We designed and performed assays for 175 of the most frequently cited human RefSeq transcripts to assess the ability of the ProbeLibrary concept to generate functional real-time gene expression assays. All assays were performed using a standard protocol, with 100 nM probe, 200 nM of each primer and Qiagen QuantiTect master mix. The assays were performed in replicates of three and were characterized as functional if the following criteria were met: (i) the real-time PCR amplification curve was sigmoid shaped, (ii) the generated fluorescence was ample, (iii) the cycle threshold (C_t) was below 37, and (iv) end-point analysis of the amplification products by agarose gel electrophoresis showed a single, correct-sized PCR amplicon. Functional assays were generated for 96% of the 175

Table 1 | Assay design success rate

First assay successful	84%
Second assay successful	12%
First and second assay unsuccessful	4%
Success rate	96%

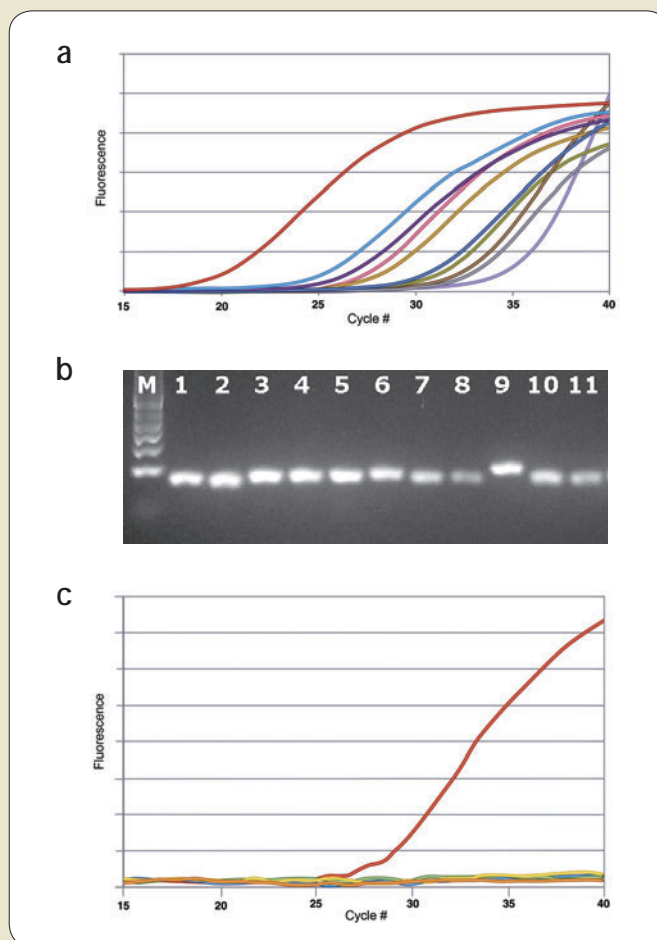


Figure 3 | Validation and performance of ProbeLibrary assays. (a) Real-time PCR amplification curves and (b) gel-electrophoresis end-point analysis of the amplification products obtained in assays designed for 175 of the most-cited RefSeq transcripts. (c) Assay specificity: the probe used in the shown assays only produced fluorescence on a perfect-matched target, whereas no signal was generated on single mismatch targets both internal 5′-1 (T:G or T:C) and 3′-1 (A:G), as well as on a dual-mismatched target.

RefSeq transcripts (**Fig. 3a,b** and **Table 1**) without any optimization of assay conditions. Assay optimization was avoided by replacing assays that performed inadequately with an alternative assay from the list provided by ProbeFinder.

The average PCR efficiency of all the functional assays was estimated to be 1.96 using LinRegPCR software². Estimation via classical calibration curve showed an average PCR efficiency of 104% using a subset (~10%) of the 175 assays. A PCR efficiency of 2.00, obtained with LinRegPCR, corresponds to 100% using calibration curves.

The specificity of ProbeLibrary assays was analyzed by real-time PCR amplification of different templates containing 0, 1 or 2 mismatches relative to the ProbeLibrary probe (**Fig. 3c**). Fluorescence was generated only when a perfect match existed in the amplicon, whereas a single mismatch even in the 3′-1 and 5′-1 position relative to the probe prevented the generation of fluorescence, indicating high probe specificity. Similar discrimination was demonstrated in real-time PCR assays using individual primer pairs from the assays designed to detect the 175

APPLICATION NOTES

most cited RefSeq transcripts, but in which the correct probe had been replaced with an alternate ProbeLibrary probe containing one or more mismatches relative to the amplified target (data not shown).

Conclusion

The ProbeLibrary concept, which consists of a collection of 90 real-time PCR probes and assay design software, combines flexibility and specificity for real-time PCR assays. Up to 99% of mRNA transcripts within one organism can be quantified with the 90 probes contained in the human, mouse, rat, Arabidopsis, Drosophila, *Caenorhabditis elegans* or Chimpanzee ProbeLibrary kits and intron-spanning assays can be designed for up to 90% of the transcripts. The ProbeFinder software provides fast and easy assay design, enabling the performance of real-time PCR assays the day after they were designed. These assays are fully compatible with existing equipment.

High coverage, easy assay design and assay compatibility provide a high degree of flexibility; at the same time, *in silico* PCR ensures high specificity of primer and probe combinations in the designed assays. This specificity is further enhanced by the use of short, LNA-spiked probes that have increased base pairing specificity to an extent that enables single-base mismatch discrimination.

The free ProbeFinder software as well as further information on the ProbeLibrary are freely available (<http://www.probelibrary.com>).

ACKNOWLEDGMENTS

Technical assistance from S. Ludvigsen, M.B. Mogensen and M. Bjørn is gratefully acknowledged.

1. Koshkin, A.A. *et al.* LNA (locked nucleic acids): synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil biocyclonucleoside monomers, oligomerisation and unprecedented nucleic acid recognition. *Tetrahedron* **54**, 3607–3630 (1998).
2. Ramakers, C., Ruijter, J.M., Deprez, R.H. & Moorman, A.F. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* **339**, 62–66 (2003).