DE GRUYTER OPEN

Editorial

DOI: 10.1515/rrlm-2015-0022

From primer design to validation of results - is it possible by using free software only?

De la proiectarea primerilor la validarea rezultatelor - este posibil utilizând doar programe gratuite?

Adrian Man[1,2*], Claudia Bănescu[3], Minodora Dobreanu[4], Cornel Fraefel[2]

1. Department of Microbiology, University of Medicine and Pharmacy Tirgu-Mures, Romania; 2. Institute of Virology, University of Zurich, Switzerland; 3. Department of genetics, University of Medicine and Pharmacy of Tîrgu Mureș, Romania; 4. Department of Laboratory Medicine, University of Medicine and Pharmacy of Tîrgu Mureș, Romania

Abstract

Successful experiments in molecular biology require good knowledge about various methods and protocols. In molecular biology, nucleic acid manipulation is the essence, starting with the quality of extraction and ending with several analysis assays (PCR, RT-PCR, qPCR, PCR arrays, molecular cloning, etc). Though many of these are so called "standardized", in practice there are many variables that can influence the outcome of the experiment. Due to the importance of optimal primer design in PCR assays, we will focus on primer designing and checking software, but we also present other useful free tools that can help researchers in the molecular biology field.

Keywords*: primer design, free software, molecular biology*

Rezumat

Pentru a avea succes în experimentele de biologie moleculară, sunt necesare cunoștințe bune despre diversele metode și protocoale în acest domeniu. În biologia moleculară, manipularea corespunzătoare a acizilor nucleici este esențială, începând cu calitatea extracției până la testele efective (PCR, RT-PCR, qPCR, PCR array, clonare, etc). Deși multe dintre protocoale sunt "standardizate", în practică sunt multe variabile care pot influența rezultatele experimentelor. Datorită importanței proiectării corecte a primerilor în metodele de tip PCR, ne vom concentra pe modul de proiectare și pe verificarea lor cu ajutorul programelor de calculator, dar de asemenea prezentăm și alte programe utile disponibile gratuit, care pot ajuta cercetătorii în domeniul biologiei moleculare.

Cuvinte cheie*: proiectare primeri, programe gratuite, biologie moleculară*

Received: 3rd May 2015; Accepted: 6th June 2015; Published: 12th June 2015

***Corresponding author:** Adrian Man, Department of Microbiology, University of Medicine and Pharmacy Tirgu-Mures, Romania; e-mail: adrian.man@rrml.ro

In PCR assays, there are a several aspects that have to be carefully optimized, and primer design is probably the most important one. Badly designed primers will fail to induce DNA amplification, or in the best of cases will lead to poor PCR efficiency and poor yield. There are several free online resources that can be used for picking PCR primers, such as **PrimerBank** provided by Harvard Medical School and Massachusetts General Hospital, which holds nearly 250.000 primer sets (1), **RTPrimerDB** provided by Ghent University, with over 8700 sets (2), or **qPrimerDepot** provided by the National Institutes of Health (3). Note that most of the listed primers are not validated, so this should be performed according to the research protocol.

One of the best known free online tools for primer design is **Primer3** developed by Howard Hughes Medical Institute or **Primer-BLAST** from National Institutes of Health. The latter uses Primer3 software to design the primers and automatically blast the results according to the required settings (4,5). Primer-BLAST provides the possibility to check primers against a custom DNA sequence (including FASTA format) or against accession or GI identification numbers from NCBI, both for DNA and mRNA (6). We have to mention that the sequences provided by NCBI as mRNA are in fact cDNA, as uracil (U) is already replaced with thymidine (T), so these sequences can be used for cDNA primer design.

The tool also provides many options that can be used for optimizing primer design. Usually, the PCR product size must not be over 200 base pairs (bp); ideally, a smaller amplicon size (~100 bp) is more adequate for qPCR, but for conventional PCR larger sizes up to 400-500 bp are also adequate. It is generally accepted that the primer length should be ~20 bp; a short primer will not have enough specificity, while a long primer will have a high melting temperature (*Tm*), especially if it contains high ratios of guanine (G) and cytosine (C). For this particular reason, the GC content of the primers should be between 40-60%; this option as well as the primer size can be set in the advanced options area of Primer-BLAST. In two-step PCR protocols, the aligning and extension steps are overlapping at the same temperature. The extension requires 72°C (the optimal temperature for Taq polymerase) for 1 minute per kb, and it will be slower at lower temperatures. Therefore, primers with higher *Tm* (62-65°C) have to be designed, and the second PCR step can be set at around 68°C. This is possible when using a single detector system such as SYBR, but not if using dual-labeled probes such as FAM-TAMRA, where the 5'-3' exonuclease activity of the Taq DNA polymerase that cleaves the probe is most active at 60°C. In three-step PCR protocols, the annealing step is conventionally set to 55°C, so the *Tm* of the primers has to be between 57-60°C (2-5°C over the annealing temperature). If in doubt, a PCR with rising annealing temperatures (e.g. 51, 53, 55, 57, 59°C) can be performed. There are useful online tools for checking *Tm* for primers including some that consider the NaCl and Mg^{2+} concentrations (7–9). Tm is increased in case of a high Mg^{2+} concentration. Usually, the final $MgCl_2$ concentration used in a standard PCR protocol is 1.5 mM. Insufficient Mg^{2+} may result in failure of the PCR reaction. In addition, the difference between the Tms of the primers should be less than 5°C.

The general formula for calculating *Tm*, which can be easily applied in spreadsheet software, is:

$$Tm\ (°C) = 81.5+16.6*log10[Na \text{ and } K \text{ in moles/L}]+0.41*(GC\%)-(\frac{675}{\text{length of primer}})$$

An easier approach for calculating the *Tm* is summing 2°C for each A or T and 4°C for each G or C in the primer sequence, but this is not recommended for sequences longer than 14 nucleotides. In case of longer primers, it is recommended to apply the following formula:

$$Tm\ (°C)=64.9+41*\frac{\text{Number of C+G-16.4}}{\text{length of primer}}$$

Besides the GC content, GC clamp can also be set, as the presence of these nucleotides at the 3' end promotes strong and specific binding of the primer, though rich GC content in one end of the primer compared to the rest will lead to poor specificity. Max Poly-X defines the maximum amount of repeated nucleotides, and has to be lower than 5, otherwise the primer can misprime.

All the primer sequences have to be blasted against the target gene (DNA or cDNA) in order to check the proper alignment, the gene specificity and the possible side products or the formation of primer dimers.

When working with cDNA, it is important not to amplify the genomic DNA. The "Exon junction span" option can be used with at least one primer spanning an exon-exon junction, thus making the primer cDNA specific. Small tunings like including at least one intron from genomic DNA between primers can help to distinguish between cDNA and genomic DNA amplicons (based on melting curve or gel electrophoresis). These settings are available only when using mRNA accession numbers starting with NM_ or genomic accession numbers starting with NG_, and not when using FASTA input sequences.

Max Self Complementarity and Max Pair Complementarity are very important settings to follow, as high values indicate high chance for secondary structure formation during the PCR run: self-dimers, hairpins and cross dimers (between primer pairs). **Beacon designer** is an online tool that gives details about possible secondary structures, the indexes being expressed as ΔG (Gibbs Free Energy G – energy required to break the secondary structure) (10). Negative values correspond to more stable structures compared to positive values. The values should be in the range of -3 to 3 kcal/mol, ideally as close to 0 as possible and in the worst case ±5. Higher values are acceptable as long as the aligning temperature in the PCR protocol is set higher. For hairpins, lower ΔG are tolerated compared to dimers. The annealing of 3' ends between primers, especially when the 3' ends are rich in GC, should be totally avoided; practically, the 3' complementarity should be 0 or 1 at maximum.

The specificity of the primers has to be considered in order to avoid their alignment with other sequences rather than the sequence of interest (genomic DNA in cDNA samples, poor designed primer). Regarding the primer design, repeated nucleotides at the primer ends and high 3' stability (high ΔG value of the last five bases from the 3' end) should be avoided.

Stable template secondary structures interfere with primer aligning and will lead to a low yield of PCR product. Thus, the primers have to be designed in order to align in regions that do not form stable secondary structures. **Predict a Secondary Structure** is a web server that computes graphs and plots on possible secondary structures on given sequences (11).

Another useful multi-purpose freeware tool is **Serial cloner**, available for all operating systems (12). With a very intuitive interface, it can help researchers to simulate PCR reactions with custom primers, to align two sequences, to design plasmids, viral vector maps and small hairpin RNA inserts, to simulate endonuclease restriction or translation, and many more. It can import several DNA and cDNA sequence formats (FASTA, GeneBank, ASCII text formats, etc) and display details regarding the GC% content, Tm, and number of nucleotides. The sequences can be virtually manipulated (create

reverse, complementary or reverse-complementary DNA strands) or blasted against the NCBI database. The software also contains basic tools like genetic code, DNA alphabet, and restriction enzyme library.

Chromas lite is a program designed for Windows operating system that can open several chromatogram file types and export the content in FASTA format (13). The chromatogram is displayed color-coded, both graphically and by nucleotide symbols. Unrecognized nucleotides can be easily identified following the corresponding chromatogram peak.

By combining these tools, researchers can easily design experiments and check their results.

List of abbreviations

cDNA - Complementary DNA
DNA - Deoxyribonucleic acid
mRNA - Messenger RNA
NCBI - National Center for Biotechnology Information
PCR - Polymerase chain reaction
qPCR - Quantitative real-time PCR
RT-PCR - Reverse transcription polymerase chain reaction

References

1. PrimerBank [Internet]. [cited 2015 May 6]. Available from: http://pga.mgh.harvard.edu/primerbank/
2. RTPrimerDB: the real-time PCR primer and probe database - Homepage [Internet]. [cited 2015 May 6]. Available from: http://medgen.ugent.be/rtprimerdb/index.php
3. qPrimerDepot [Internet]. [cited 2015 May 6]. Available from: http://primerdepot.nci.nih.gov/
4. Primer designing tool [Internet]. [cited 2015 May 6]. Available from: http://www.ncbi.nlm.nih.gov/tools/primer-blast/
5. Primer3 Input [Internet]. [cited 2015 May 6]. Available from: http://biotools.umassmed.edu/bioapps/primer3_www.cgi
6. National Center for Biotechnology Information [Internet]. [cited 2015 May 6]. Available from: http://www.ncbi.nlm.nih.gov/
7. Tm Calculator | Life Technologies [Internet]. [cited 2015 May 30]. Available from: https://www.lifetechnologies.com/ch/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html
8. Melting Temperature (Tm) Calculation [Internet]. [cited 2015 May 6]. Available from: http://www.biophp.org/minitools/melting_temperature/demo.php
9. Tm Calculator [Internet]. [cited 2015 May 30]. Available from: http://appliedbiosystem.com/support/techtools/calc/
10. Beacon Designer Free Edition [Internet]. [cited 2015 May 6]. Available from: http://www.premierbiosoft.com/qOligo/Oligo.jsp?PID=1
11. Mathews Lab RNAstructure Web Servers [Internet]. [cited 2015 May 6]. Available from: http://rna.urmc.rochester.edu/RNAstructureWeb/index.html
12. Serial Cloner [Internet]. [cited 2015 May 7]. Available from: http://serialbasics.free.fr/Serial_Cloner.html
13. Technelysium | Chromas [Internet]. [cited 2015 May 7]. Available from: http://technelysium.com.au/?page_id=13