BIOINFORMATICS APPLICATIONS NOTE

Vol. 25 no. 2 2009, pages 276–278
doi:10.1093/bioinformatics/btn614

Gene expression

# MFEprimer: multiple factor evaluation of the specificity of PCR primers

Wubin Qu[1,2], Zhiyong Shen[1], Dongsheng Zhao[3], Yi Yang[2,*] and Chenggang Zhang[1,*]

[1]Beijing Institute of Radiation Medicine, State Key Laboratory of Proteomics, Beijing 100850, [2]College of Life Science, Sichuan University, Key Laboratory of Bio-Resources and Eco-Environment of MOE, Chengdu 610064 and [3]Beijing Institute of Health Service and Medical Information, Beijing 100850, China

Received on August 17, 2008; revised on October 26, 2008; accepted on November 23, 2008

Advance Access publication November 27, 2008

Associate Editor: David Rocke

**ABSTRACT**

**Summary:** We developed a program named MFEprimer for evaluating the specificity of PCR primers based on multiple factors, including sequence similarity, stability at the 3′-end of the primer, melting temperature, GC content and number of binding sites between the primer and DNA templates. MFEprimer can help the user to select more suitable primers before running either standard or multiplex PCR reactions. The cDNA and genomic DNA databases of 10 widely used species, as well as user custom databases, were used as DNA templates for analyzing primers specificity. Furthermore, we maintained a Primer3Plus server with a modified Primer3Manager for one-stop primer design and specificity checking.

**Availability:** The web service of MFEprimer is freely available at http://biocompute.bmi.ac.cn/MFEprimer/. A stand-alone version of MFEprimer can be downloaded at http://biocompute.bmi.ac.cn/MFEprimer/download/. Primer3Plus on our server is located at http://biocompute.bmi.ac.cn/MFEprimer/primer3plus.cgi. All programs are licensed under the General Public License (Free Software Foundation, 1991).

**Contact:** zhangcg@bmi.ac.cn

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

## 1 INTRODUCTION

Specific primer set is crucial for a successful PCR reaction. Although a large number of PCR primer design programs have been developed, few are available for the user to evaluate primer specificity. Both PerlPrimer (Marshall, 2004) and Primer3Plus (Untergasser *et al*., 2007) only provide a link to the NCBI BLAST homepage for the user to manually check the specificity of their primers using public cDNA or genomic DNA sequence databases. Other programs, such as VPCR (Lexa *et al*., 2001), PRIMEX (Lexa and Valle, 2003), PUNS (Boutros and Okey, 2004) and GenomeTester (Andreson *et al*., 2007), focus on the sequence similarity analysis between primers and the DNA templates, despite the well-documented importance of other factors necessary for a successful PCR reaction. For example, stability at the 3′-end of the primer is the key factor for specific elongation in PCR reactions (Miura *et al*., 2005; Onodera and Melcher, 2004), and it also contributes substantially to nonspecific extension in PCR reactions, especially when the binding site at this position is relatively stable. Accordingly, the Gibbs free energy ($\Delta G$) for the last five residues at the 3′-end should be higher than $-9$ kcal/mol (Rychlik, 1995; Wang and Seed, 2003). At the same time, melting temperature ($T_m$) is one of most important characteristics affecting the efficiency and accuracy of PCR reactions (Chavali *et al*., 2005). The preferred range of $T_m$, 50–80°C, is usually considered by popular primer design programs (Rozen and Skaletsky, 2000) and primer databases (Han and Kim, 2006; Wang and Seed, 2003). In addition, the number of binding sites (the less, the better), GC content (the more close to 50%, the better) and number of predicted PCR products (the less, the better) are major factors that can cause PCR reaction failure when the DNA template is a large genome, for example the human genome (Andreson *et al*., 2008). Here, we introduce a new PCR primer specificity analysis program named MFEprimer to solve this problem. MFEprimer considers not only the sequence similarity between primers and DNA templates, but also the abovementioned factors critical to successful PCR reactions. Databases used by the program for specificity checking are comprehensive, and approximately 10 widely used species are available for *in silico* analysis. The input/output interface is user friendly and well-organized for easily understanding the specificity of each primer set, together with the sequence of the predicted PCR product(s). MFEprimer can be used to check the specificity of primer set(s) for both standard and multiplex PCR to save cost and time and thus in turn improve the efficiency of PCR reactions. We provide both the web-based and stand-alone version of this tool for different purposes. The web service is convenient with a friendly interface, while the stand-alone version is suitable for high-throughout specificity checking of primers and for customization of the databases in order to meet specific user demands. For example, if the user wish to check the specificity of a primer set for amplifying the highly conserved genes, such as $\beta$-actin in human, rat and mouse, the user may run the stand-alone version of MFEprimer by combining these databases together to complete the evaluation. In addition, we maintained a Primer3Plus server with a modified Primer3Manager, which bridges Primer3Plus (Untergasser *et al*., 2007) and MFEprimer together for one-stop service of primer design and specificity checking.

*To whom correspondence should be addressed.

 © The Author 2008. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

## 2 ALGORITHM AND IMPLEMENTATION

MFEprimer uses NCBI BLASTN (Altschul *et al*., 1990) (version 2.2.18) to predict the potential PCR products by primer similarity analysis. Mismatches are allowed except for the last three residues of the 3′-end of the primer (Sommer and Tautz, 1989). Default value of the major parameters for running NCBI BLASTN are $e = 1000$ for (expectation value) and $W = 11$ (word size). The cDNA and genomic DNA databases of 10 widely used species (*Homo sapiens, Rattus norvegicus, Mus musculus, Gallus gallus, Danio rerio, Drosophila melanogaster, Caenorhabditis elegans, Arabidopsis thaliana, Saccharomyces cerevisiae* and *Escherichia coli*) were downloaded from NCBI (ftp://ftp.ncbi.nih.gov) and are updated quarterly each year. The RefSeq database of these species are mounted as mRNA/cDNA templates for checking primer specificity, and genomic DNA databases (reference assembly version) can be used for predicting PCR product(s) and evaluating the performance of PCR when using chromosomal DNA as the template. The mitochondrial genomic DNA sequences are merged into the genomic DNA databases of these species for convenient analysis, because they are usually present in genomic DNA samples. Besides, user custom databases are also supported by MFEprimer.

MFEprimer is easily understandable, and takes into consideration of the most important factors affecting PCR performance, such as $\Delta G$ and $T_m$, to help the user have a detailed understanding of the primer's specificity. The nearest neighbor model with salt correction is used to calculate the $T_m$-value (SantaLucia, 1998; von Ahsen *et al*., 2001), while the thermodynamic parameters are obtained from the literature (SantaLucia, 1998). Mismatched base-pairing is also taken into account and the thermodynamic parameters are taken from the literature (Allawi and SantaLucia, 1997, 1998a, b, c; Peyret *et al*., 1999). The number of binding sites, GC content and the number of predicted PCR products of these primers are also output in MFEprimer (Andreson *et al*., 2008). In addition, a new parameter of primer pair coverage (PPC) is introduced to score the ability of the primer pair to bind to the DNA template (see formula below).

$$\mathrm{PPC} = \frac{Fm}{Fl} \times \frac{Rm}{Rl} \times CVfr$$

Where *Fm* and *Rm* are sequence overlaps of the forward primer and reverse primer with the DNA template, and *Fl* and *Rl* are the full lengths of the forward primer and reverse primer, respectively. *CVfr* is the coefficient of variability of matched length of the *Fm* and *Rm*. The maximum value of PPC is 100%, indicating that the primer pair can bind to the template over their entire lengths.

The function 'Filter Results' is introduced in MFEprimer to filter the amplicons unsatisfied to user-defined PCR conditions. The filters included $\Delta G$, $T_m$, PPC and amplicon size. For example, if the $T_m$ is set between 50°C and 80°C, the predicted amplicon with a $T_m$ out of this range will be eliminated. In addition, a primer set should also be discarded if they result in nonspecific amplicon(s) with similar or close size to that of the specific amplicon. However, if the size of the specific amplicon and nonspecific ones can be easily separated, for example by 1–2% agarose gel, the primer set is still acceptable.

The web server for MFEprimer is written in CGI/Python. A stand-alone version of MFEprimer is a command line program written in Python and can be executed on any UNIX-like systems. All the programs described here are licensed under the General Public License (Free Software Foundation, 1991).

## 3 DISCUSSION

NCBI BLAST is one of the fastest local alignment search programs for checking primer specificity. Other tools, such as BLAT (Kent, 2002), may be faster than NCBI BLAST in searching very large databases, such as the human genome, with long queries. However, BLAT is not suitable for analyzing the primer specificity, which are typically shorter than 30 bp and usually have a word size shorter than 8. We have compared NCBI BLAST (version 2.2.15 and 2.2.18) with other similarity search programs such as MPBLAST (Korf and Gish, 2000) and miBLAST (Kim *et al*., 2005), and found that NCBI BLAST version 2.2.18 is the fastest program (data not shown). As the parameters *e* (expectation value: for a given score, the number of hits in a database search that was expected to see by chance with this score or better) and *W* (word size: the size of the initial word that must be matched between the database and the query sequence) used in NCBI BLAST have significant effects on the sensitivity and speed of MFEprimer, we suggest to use $e = 1000$ and $W = 11$ for balancing the sensitivity and speed performance of the program. Not surprisingly, the parameters of $e > 1000$ and $W < 11$ allow more stringent analysis of primer specificity, but they require longer computation time.

However, MFEprimer does not support degenerate primers currently, because the NCBI BLAST treats degenerate base as mismatch in the alignment and in turn preventing BLAST to find any matches in the database. Therefore, MFEprimer does not work for degenerate primers and will be modified in later version.

## ACKNOWLEDGEMENTS

The authors wish to thank a lot to the anonymous reviewers for their valuable suggestions to this article.

*Funding*: National Basic Research Project (973 program) (2003CB715900, 2006CB504100, 2006CB0D0807); General Program (30771230, 30600664) of the Natural Science Foundation of China; Major Program for Science and Technology Research of Beijing Municipal Bureau (7061004).

*Conflict of Interest*: none declared.

## REFERENCES

Allawi,H.T. and SantaLucia,J. (1997) Thermodynamics and NMR of internal GT mismatches in DNA. *Biochemistry*, **36**, 10581–10594.

Allawi,H.T. and SantaLucia,J. (1998a) Nearest-neighbor thermodynamics of internal A center dot C mismatches in DNA: sequence dependence and pH effects. *Biochemistry*, **37**, 9435–9444.

Allawi,H.T. and SantaLucia,J. (1998b) Nearest neighbor thermodynamic parameters for internal G center dot A mismatches in DNA. *Biochemistry*, **37**, 2170–2179.

Allawi,H.T. and Santalucia,J. (1998c) Thermodynamics of internal C center dot T mismatches in DNA. *Nucleic Acids Res.*, **26**, 2694–2701.

Altschul,S.F. *et al*. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Andreson,R. *et al*. (2007) Fast masking of repeated primer binding sites in eukaryotic genomes. *Methods Mol. Biol.*, **402**, 201–217.

Andreson,R. *et al*. (2008) Predicting failure rate of PCR in large genomes. *Nucleic Acids Res.*, **36**, e66.

Boutros,P.C. and Okey,A.B. (2004) PUNS: transcriptomic- and genomic-in silico PCR for enhanced primer design. *Bioinformatics*, **20**, 2399–2400.

Chavali,S. *et al*. (2005) Oligonucleotide properties determination and primer designing: a critical examination of predictions. *Bioinformatics*, **21**, 3918–3925.

Han,S. and Kim,D. (2006) AtRTPrimer: database for Arabidopsis genome-wide homogeneous and specific RT-PCR primer-pairs. *BMC Bioinformatics*, **7**, 179.

Kent,W.J. (2002) BLAT–the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.

Kim,Y.J. *et al*. (2005) miBLAST: scalable evaluation of a batch of nucleotide sequence queries with BLAST. *Nucleic Acids Res.*, **33**, 4335–4344.

Korf,I. and Gish,W. (2000) MPBLAST: improved BLAST performance with multiplexed queries. *Bioinformatics*, **16**, 1052–1053.

Lexa,M. *et al*. (2001) Virtual PCR. *Bioinformatics*, **17**, 192–193.

Lexa,M. and Valle,G. (2003) PRIMEX: rapid identification of oligonucleotide matches in whole genomes. *Bioinformatics*, **19**, 2486–2488.

Marshall,O.J. (2004) PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics*, **20**, 2471–2472.

Miura,F. *et al*. (2005) A novel strategy to design highly specific PCR primers based on the stability and uniqueness of 3'-end subsequences. *Bioinformatics*, **21**, 4363–4370.

Onodera,K. and Melcher,U. (2004) Selection for 3' end triplets for polymerase chain reaction primers. *Mol. Cell. Probes*, **18**, 369–372.

Peyret,N. *et al*. (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A center dot A, C center dot C, G center dot G, and T center dot T mismatches. *Biochemistry*, **38**, 3468–3477.

Rozen,S. and Skaletsky,H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365–386.

Rychlik,W. (1995) Priming efficiency in PCR. *Biotechniques*, **86**, 88–90.

SantaLucia,J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.

Sommer,R. and Tautz,D. (1989) Minimal homology requirements for PCR primers. *Nucleic Acids Res.*, **17**, 6749–6749.

Untergasser,A. *et al*. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.*, **35**, W71–W74.

von Ahsen,N. *et al*. (2001) Oligonucleotide melting temperatures under PCR conditions: nearest-neighbor corrections for Mg2+, deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas. *Clin. Chem.*, **47**, 1956–1961.

Wang,X.W. and Seed,B. (2003) A PCR primer bank for quantitative gene expression analysis. *Nucleic Acids Res.*, **31**, e154.