



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Selection for 3' end triplets for polymerase chain reaction primers

Kenji Onodera¹, Ulrich Melcher*

Department of Biochemistry and Molecular Biology, Oklahoma State University, 246 NRC, Stillwater, OK 74078, USA

Received 23 March 2004; accepted for publication 28 May 2004

Abstract

The 3' end of a primer is a key component of PCR primer design. Many recommendations for the composition and sequence of the 3' end have been suggested based on theoretical considerations, but have not been verified experimentally. We analyzed 3' end triplets of PCR primer sequences obtained from refereed journal articles, to test those recommendations and to make empirical recommendations for primer design. The frequencies of the 64 possible 3' end triplets among 2137 PCR primers from the VirOligo database were not uniformly distributed. From the analysis, we found that unfavored and preferred 3' end triplets existed, and that the apparent preferences were not due to base compositions in viral genome sequences. Comparison of the sequences preferred by practitioners to those recommended, suggested that no single recommendation is entirely satisfactory. We suggest that recommendations be replaced with a scoring system incorporating empirical frequencies such as those reported here.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: PCR; Primer design; Oligonucleotides; Viruses; VirOligo

1. Introduction

Many current molecular biological analyses, especially microarray-based analysis of gene expression, require the efficient design of large numbers of polymerase chain reactions (PCRs). In the PCR, the 3' ends of primers must anneal to templates to be elongated by a DNA polymerase. A high G+C content at the 3' end of a primer may obviate the need for complete complementarity and annealing of the remainder of the primer sequence, thus diminishing the specificity of the priming reaction [1]. A low G+C content at a primer's 3' end may, due to weak annealing, increase the importance of complete annealing of the remainder of the primer sequence [2]. On the other hand, such primers may not be elongated efficiently by DNA polymerase [1]. Although the importance of the 3' end triplet in primer design is recognized, many recommendations for 3' end

triplet selection, based mostly on theory, have been put forth, and some contradict others. Recommended compositions of 3' ends are: no T at the 3' end and at least one W (A or T) in the 3' end triplets [3]; S (C or G) at the 3' end and no GC or CG, due to potential formation of hairpins and primer-dimers [4]; low G+C [2,5]; one or two S [6]. One way to test the validity of those recommendations and determine which 3' end triplets are best suited for PCR primers is to examine the 3' ends of primers actually used in successful PCR experiments. We here report such an examination.

The VirOligo database [7] contains a large number of oligonucleotide sequences used in the detection of viruses, the majority of detection assays being by PCR. These sequences were used in actual successful PCR experiments and published in refereed journals. The primers and hybridization probes, along with the experimental conditions for their use, are publicly available at <http://VirOligo.okstate.edu/> [7]. Only PCR primer sequences in the VirOligo database were obtained for the analysis. The analysis results revealed preferred and disfavored 3' end triplets in successful primer sequences, and led to recommendations for primer design based on experience rather than theory.

* Corresponding author. Tel.: +1-405-744-6210; fax: +1-405-744-7799.

E-mail addresses: onodera@gsc.riken.jp (K. Onodera), u-melcher-4@alumni.uchicago.edu (U. Melcher).

¹ Present address: Genomic Sciences Center, Riken, Yokohama 230-0045, Japan.

2. Materials and methods

Primer sequences and PCR conditions were obtained for the analysis from the VirOligo database on February 6, 2002. On that date, the VirOligo database covered a total of 1685 articles whose abstracts appeared in PubMed before December 19, 2001. The articles contained 3985 published virus-specific oligonucleotide sequences and 2300 PCR and hybridization conditions for detection of the viruses for alcelaphine herpesvirus, bovine adenovirus, bovine viral diarrhea virus (BVDV), bovine herpesvirus (BHV), bovine respiratory syncytial virus, bovine rotavirus, bovine coronavirus, foot-and-mouth disease virus (FMV), variola (smallpox) virus, cowpox virus, and human adenovirus. Duplicated primer sequences were eliminated from the analyzed set. The frequencies of occurrence of each of the 64 possible 3' end triplets, defined as the number of occurrences of a triplet divided by the total number of primers $\times 100$ (%), were calculated. For selected viruses, the frequency distributions of overlapping triplets in the genome sequences were calculated for comparison with the triplet frequencies of the 3' ends of primers specific for those viruses. The viruses were BHV-1 (GenBank Accession No. AJ004801; genome size 135 kb), BVDV (AF220247; 12 kb) and FMV (AF377945; 8 kb). Statistical analyses were performed by Analysis ToolPak provided by Microsoft Excel 2000.

3. Results

The 3' end triplets were obtained from 2137 virus-specific PCR primers registered in the VirOligo database. Overall, the average frequencies of triplets in the four base composition classes W_3 , W_2S , WS_2 and S_3 were, respectively, 0.91, 1.50, 1.93, and 1.30%. The frequencies of the 64 triplets were not uniformly distributed (Fig. 1; completely random distribution rejected by chi-square test, $P < 0.001$). The mean and standard deviation (s.d.) of the distribution were 1.56 and 0.63%, respectively. The most popular triplet, AGG (3.27%), was 7.8 times more frequently used than the least popular triplet, TTA (0.42%). Most frequently reported triplets (frequencies greater than mean plus s.d.) were of eleven types, AGG (3.27%), TGG (2.95%), CTG (2.85%), TCC (2.76%), ACC (2.76%), CAG (2.71%), AGC (2.57%), TTC (2.48%), GTG (2.48%), CAC (2.38%), and TGC (2.34%). All but one of the eleven triplets (TTC) had two S and one W. Six of these 11 triplets were WSS sequences. However, two of eight WSS triplets, ACG (1.31%) and TCG (1.08%), were two to three times less frequently reported in the VirOligo database than other combinations of WSS. WGC was popular and in the top 11 combinations, but the proportion of WCG was less than the average. Four of the top 11 triplets were SWS triplets. All top 16 reported triplets had an S at their 3' ends. TTS was also frequently reported, and the two TTS triplets were among the top 14.

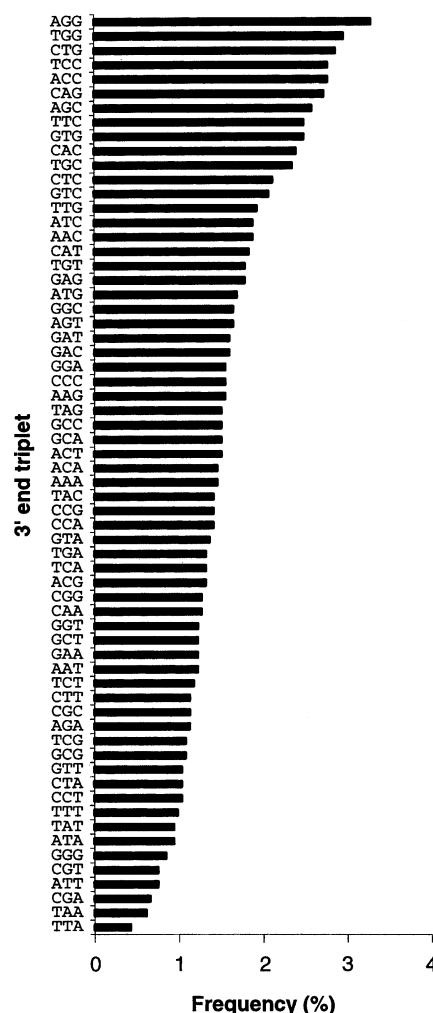


Fig. 1. Distribution of the 3' end triplets of PCR primers.

The least frequently reported triplets (less than mean minus s.d.) were of six types, TTA (0.42%), TAA (0.61%), CGA (0.65%), ATT (0.75%), CGT (0.75%), and GGG (0.84%). Two of the lowest six types contained CGW. As mentioned above, WCG was less frequently reported compared to other combinations of WSS. Three out of the lowest six types were triplets devoid of G or C, WWW. Even the most frequently reported types in WWW combinations, AAA and AAT, were distributed less frequently than average (1.45 and 1.22%, respectively).

We considered the possibility that the distribution of primer 3' triplet frequencies represents the distribution of triplets in the target viral genome sequences rather than any PCR-driven preference. Since the primer sequence data set contained 134 BHV-1, 237 BVDV and 121 FMV primers, these viruses were used to test whether genome compositions determined the 3' end triplet frequencies. Thus, triplet frequencies of the BHV-1 genome (135,382 overlapping triplets) were subtracted from frequencies of 3' end triplets in BHV specific primers (Fig. 2) to yield differential frequencies. The same calculation was performed for BVDV and FMV (12,274 and 7794 overlapping triplets,

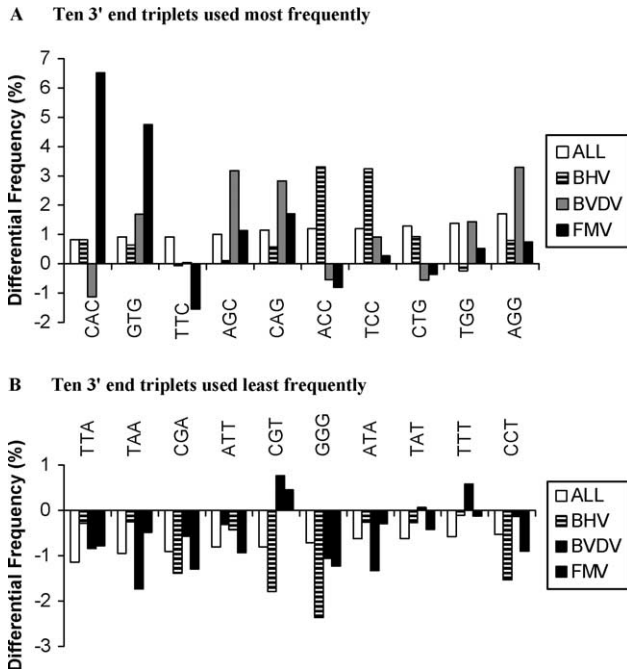


Fig. 2. Comparison among differential triplet frequencies of all primers (ALL) and BHV, BVDV, and FMV specific primers. The differential 3' end triplet frequency values for all primers represent 3' end triplet frequencies (%) minus average, and show differences from random distributions of triplets. Values for BHV, BVDV and FMV specific primers represent 3' end triplet frequencies (%) minus triplet frequencies (%) in the genome sequences, and show differences from their genome compositions. Negative values in BHV, BVDV and FMV indicate under-representation of the triplets and positive values indicate over-representation of the triplets.

respectively). In this analysis, a differential frequency of zero would indicate that a primer's triplet frequency is expected from its representation in the genome sequence. Negative values indicate that those triplets were used less frequently in primers than in genome sequences, while positive values indicate more frequent use.

Differential frequencies for the 10 most and least frequently used 3' end triplets from Fig. 1 are shown in Fig. 2 for BHV-1, BVDV and FMV, along with the differences of the triplet frequencies for all primers analyzed from the mean frequencies for those primers. Among the 30 comparisons, 22 of the 10 most frequently used 3' end triplets were over-represented compared to the genome sequences (Fig. 2A). For the 10 least frequently used 3' end triplets, 26 of 30 comparisons revealed under-representation (Fig. 2B). Means of differential frequencies between top ten and least ten 3' end triplets were significantly different for all three viruses tested (t test, $p=0.011$ for BHV; $p=0.008$ for BVDV; $p=0.047$ for FMV). Since the number of primer sequences for the individual viruses were only two to four times the number of possible triplets, individual differential frequency values cannot be compared statistically. Nevertheless, the difference between the means suggests that the trends in 3' end triplet frequencies (Fig. 1) were not due to genome compositions.

4. Discussion and conclusions

Proper primer design is a key for successful PCR. The analysis of 3' end triplets used in successful PCRs (Fig. 1) revealed that these triplets were not distributed equally, suggesting that the sequence of the 3' end triplet affects PCR efficiency. Focused experiments comprehensively varying 3' end triplets of primers at a single template site could potentially reveal principles for the choice of 3' end triplets for the design of efficient primers. However, the outcome of such experiments will likely depend on the sequence context of the varying triplet and on conditions of the PCR experiment. Thus, we adopted the approach of examining primers used in successful PCRs with a variety of templates and under a variety of conditions [8].

This approach has two possible difficulties. The first, that the frequency of a 3' end triplet may predominantly be determined by its frequency in the target genome sequence, was discounted by comparison of triplet frequencies in the genomes of three viruses with the 3' end triplet frequencies (Fig. 2). The second is that the observed distribution of frequencies may be a direct result of recommendations made for primer design in the literature or as part of primer design applications. For three of the four recommendations listed in the introduction, all but AGC of the top ten 3' end triplets conform to the recommendations and AGC only fails one of the three tests (no GC). However, of the least frequently used ten 3' end triplets, only one (ATT) does not conform to at least one of these recommendations. The fourth recommendation, that the 3' end be low in %GC was violated by all but one of the most frequently used triplets, while more than half of the least frequently used ones obeyed this recommendation.

It is likely that primer design programs significantly affected the observed 3' end triplet distribution. OMIGA (Accelrys, San Diego) searches for primers with WSS at the 3' end as the default setting suggesting that all combinations of WSS triplets might be selected at the same rate. However, two of eight WSS triplets, ACG and TCG (WCG), were in the bottom half of triplet frequencies seen less frequently in the VirOligo database while the other combinations of WSS appeared in the top dozen. Other applications, such as DS genes (Accelrys), allow the user to choose the 3' end sequence but have no 3' end selection in their default settings. According to O'Connell [5], the current trend in programming choices for primers is to ask for low %GC toward the 3' end. As mentioned above, the VirOligo distribution is not consistent with this recommendation. Thus, it is unlikely that the frequencies of 3' end triplet observed in primers (Fig. 1) are due primarily to primer design software or literature recommendations.

The 3' end triplet sequence of PCR primers is one of the important factors for primer design. In the VirOligo database, WSS (19.0%) and SWS (18.0%) were the two most common 3' end triplet combinations (Fig. 1). The under-representation of WCG, mentioned above, suggests

a high frequency of failure of PCRs using primers ending in WCG. Low G+C contents in the 3' end triplets of PCR primers were not frequently reported in the VirOligo database. Among the existing recommendations, the best performer, when judged by the observed 3' end triplet frequencies is the one that requires an S at the 3' end and no GC or CG in the sequence. All but one (AGC) of the top 10 conform to this recommendation and only one (GGG) of the 10 least frequent conforms to the recommendation. Based on our results we would recommend WSS or SWS 3' ends, but avoiding WCG. However, there are additional exceptions to even this recommendation. We suggest that primer design software incorporate scores based on empirical frequencies of 3' end triplets, such as those reported here, into the evaluation of oligonucleotides as primers.

Improvement of the PCR primer success ratio will reduce reagent costs and labor in large scale efforts such as the amplification of genes or gene fragments for preparation of microarrays for gene expression analysis. It should also be highly important for efforts to design microarrays on which immobilized PCR reactions can occur, such as that proposed for the detection of viruses in virus signature amplification, or ViSA card [9].

Acknowledgements

The authors acknowledge the support of the Samuel Roberts Noble Foundation, the Robert J. Sirny Professorship (to UM) and the Oklahoma Agricultural Experiment

Station whose director has approved the manuscript for publication. The authors thank Jean d'Offay, Richard Essenberg and Robert Matts for critical review of the manuscript before submission.

References

- [1] Mitsuhashi M. Technical report: Part 2. Basic requirements for designing optimal PCR primers. *J Clin Lab Anal* 1996;10:285–93.
- [2] Hyndman D, Cooper A, Pruzinsky S, Coad D, Mitsuhashi M. Software to determine optimal oligonucleotide sequences based on hybridization simulation data. *BioTechniques* 1996;20:1090–4. See also pages 1096–7.
- [3] Kidd KK, Ruano G. Optimizing PCR. In: McPherson MJ, Hames BD, Taylor GR, editors. *PCR 2: a practical approach*. Oxford: IRL Press; 1995, p. 1–21.
- [4] Sambrook J, Russell DW. In vitro amplification of DNA by the polymerase chain reaction. In: Sambrook J, Russell DW, editors. *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor Laboratory; 2001, p. 81–817.
- [5] O'Connell J. The basics of RT-PCR. In: O'Connell J, editor. *RT PCR protocols*, vol. 193. Totowa, NJ: Humana Press; 2002, p. 19–25.
- [6] Sharrocks AD. The design of primers for PCR. In: Griffin HG, Griffin AM, editors. *PCR technology: current innovations*. Boca Raton, FL: CRC Press; 1994, p. 5–12.
- [7] Onodera K, Melcher U. VirOligo: a database of virus-specific oligonucleotides. *Nucleic Acids Res* 2002;30:203–4.
- [8] Onodera K. Construction, application and analysis of the oligonucleotide database, VirOligo, PhD thesis, Oklahoma State University, Stillwater, 2002.
- [9] Onodera K, d'Offay J, Melcher U. Nylon membrane-immobilized PCR for detection of bovine viruses. *BioTechniques* 2002;32:74–6. See also pages 78, 80.