PCR primer selection tool optimized for high-throughput proteomics and structural genomics

Jaume M. Canaves, Andrew Morse, and Bill West
University of California San Diego, La Jolla, CA, USA

BioTechniques 36:1040-1042 (June 2004)

The increasing availability of genomic data is providing the scientific community with an unprecedented amount of information. Consequently, the elucidation of the roles of genes and proteins requires the implementation of novel high-throughput approaches. One of the goals of the Joint Center for Structural Genomics (JCSG) (1), which is one of nine pilot centers funded by the Protein Structure Initiative (2-4), is the development of high-throughput methods for large-scale protein structure production. The Primer Selection Tool presented here is a JCSG bioinformatics tool (5) capable of designing large sets of oligonucleotide primers for the simultaneous PCR amplification of full-length open reading frames (ORFs) under the same experimental conditions. To the best of our knowledge, only two similar systems are currently available: (i) the Java®-based Express Primer Tool designed by the Midwest Center for Structural Genomics (6), and (ii) Xpression Primer™ (PREMIER Biosoft International, Palo Alto, CA, USA), a commercial software product for the Windows® and Mac® platforms.

The JCSG Primer Selection Tool has been implemented as a web-based application, providing multi-user capability and easy access. The current implementation of the Tool can generate 1000 primer pairs in <1 min, including input file upload through a broadband Internet connection. There is no limit in the number of sequences or sequence length that can be uploaded. The Tool is accessible through the JCSG web site (http://www.jcsg.org) under Links > JCSG Tools > Primer Selection Tool. The engine underlying the Primer Selection Tool is a Perl module freely available on the BioTechniques’ web site at http://www.BioTechniques.com/June2004/CanavesSoftware.html. The Primer Selection Tool writes the oligonucleotide primer sequences in 5’ to 3’ orientation.

After uploading the DNA sequence data and entering all the optional parameters, a collection of primers with $T_m$’s within the user-defined tolerance ranges is calculated for each target sequence. If the calculated primers are not within the user-defined length constraints, primers are discarded. Because of the speed of the program, introducing minor changes in $T_m$ tolerances, primer lengths, and re-calculating the primers is a fast and effective way to force the identification of acceptable primers.

The $3’$ terminal position in primers is essential for the control of mispriming. The presence of G or C bases at the $3’$ end of primers (GC clamp) helps to promote correct binding due to the stronger hydrogen bonding of G and C bases. GC clamp scores are calculated based on the 3 last bases of each primer according to the following schema: [GC][GC][GC] = 0; [ATGC][ATGC][AT] = 1; [ATGC][AT][GC] = 2; and [AT][GC][GC] = 3, with 0 corresponding to the worst GC clamp and 3 corresponding to the best, respectively. The primer selected for each target corresponds to the primer within the selected $T_m$ range that has best GC clamp.

The program not only calculates and outputs the $T_m$’s for the selected primers, but also their sequences, lengths, GC clamp scores, and GC content in percentage value (Figure 1B). Ideally, the GC content of primers should be between 45% and 55%, although we have not experienced PCR failures due to lower GC contents as long as the primers have optimal GC clamps.
Therefore, GC content is not used as a filtering criterion and acceptance or rejection of primers based on their GC content is left to the user.

The Primer Selection Tool has been successfully used for large-scale design of both prokaryotic and eukaryotic primers. Table 1 reports the success rates for a representative set of 874 eukaryotic and prokaryotic targets processed in 96-sample batches in a plate-based automated high-throughput setup. In two independent experiments (93% and 98% success rates), 362 of 380 eukaryotic proteins were successfully amplified. Comparable high-throughput experiments with primers designed using the Midwest Center for Structural Genomics’ Express Primer Tool (6) show amplification success rates for single organism plates between 72% and 88%. In another three independent experiments, we amplified 494 prokaryotic targets from bacterial and archaeal genomes, with success rates ranging from 60% to 94%. The differences in success correlated with the number of template sources used in each experiment (Table 1). Larger pools of species result in primers with wider ranges of GC content, Tm, and primer lengths, possibly causing a drop in PCR efficiency with respect to experiments with more homogenous PCR templates and primers. Although the efficiencies in our complex plates containing targets from six species are 85%–94%, the JCSG Primer Selection Tool still surpasses the performance of the Express Primer Tool. Only for a highly complex plate (11 bacterial and archaeal species per 96-well plate) does our efficiency drop below that achieved by the Express Primer Tool.

In summary, we present a new primer design tool specifically designed for high-throughput proteomics or genomics pipelines. The performance of the tool compares very favorably to the other web-based program currently available, namely Express Primer Tool (6), both in success rate and speed. Also, our product compares favorably to Xpression Primer regarding speed (300 sequences in 2 min for Xpression Primer), cost, and platform independence. We show that based on well-known and sound primer design concepts (8,9), it is possible to implement

<table>
<thead>
<tr>
<th>Source</th>
<th>Experiment (No.)</th>
<th>Species/96-Well Plate</th>
<th>Targets</th>
<th>Amplification Success Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eukaryotic</td>
<td>1</td>
<td>1</td>
<td>190</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>190</td>
<td>93</td>
</tr>
<tr>
<td>Prokaryotic</td>
<td>3</td>
<td>6b</td>
<td>115</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6c</td>
<td>189</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>11d</td>
<td>190</td>
<td>60</td>
</tr>
</tbody>
</table>

Genomic DNAs from one eukaryote, the yeast Saccharomyces cerevisiae, and 13 prokaryotes (four Archaea and nine Bacteria) to be used as templates for PCR amplification were obtained from ATCC (Manassas, VA, USA). Oligonucleotide primers were obtained from Operon® (Qiagen, Valencia, CA, USA). The full-length open reading frames targeted for amplification were nonredundant (no pair of protein sequences shared over 90% identity). Touchdown PCR amplification was performed in a DNA Engine Tetrad™ PTC-225™ Thermal Cycler (MJ Research, Waltham, MA, USA), using 96-well plates and 100-μL reaction volumes with kinased primers. Plates were sealed with rubber mats (USA Scientific, Ocala, FL, USA). Further detail on PCR conditions is available as Supplementary material on the BioTechniques’ web site at http://www.biotecniques.com/June2004/CanavesSupplementary.htm. Detailed information about the targets can be accessed at the Joint Center for Structural Genomics (JCSG) web site at http://www.jcsg.org.

Figure 1. The public interface of the JCSG Primer Generation Tool. (A) View of the tool’s data input user interface. (B) Summary screen and pop-up window with output files. Independently of the output format options selected by the user, the program outputs a summary page with the parameters used for the calculations as well as the number of successes and failures. The pop-up window contains the output files, including raw data, the user-selected output files, and exception/failure logs. Separate forms are produced for forward and reverse primers. JCSG, Joint Center for Structural Genomics.
a simple albeit highly efficient tool for the design of a large set of primers for plate-based high-throughput experiments. The existence of this tool, coupled with robotics, is important for the improvement in productivity of high-throughput genomics and contributes to further our knowledge of protein structure and function by accelerating the exploration of the increasingly vast amounts of genomic data.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIH) Protein Structure Initiative grant no. P50-GM 62411 from the National Institute of General Medical Sciences (http://www.nigms.nih.gov). We thank the rest of team members at the JCSG Bioinformatics Core for their intellectual and material support and helpful discussions. We also thank Rebecca Page and Jeffrey Velasquez from the JCSG Crystallography Core at The Scripps Research Institute (La Jolla, CA) for help with the preparation of the manuscript.

REFERENCES


Received 22 January 2004; accepted 7 April 2004.

Address correspondence to Jaume M. Canaves, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92037-0527, USA. e-mail: jcanaves@ucsd.edu