



AcePrimer: automation of PCR primer design based on gene structure

Sheldon J. McKay and Steven J.M. Jones*

Genome Sciences Centre, British Columbia Cancer Agency, 600 W 10th Avenue, Vancouver, BC, Canada V5Z 4E6

Received on December 4, 2001; revised on February 12, 2002; April 22, 2002; accepted on April 30, 2002

ABSTRACT

Summary: AcePrimer is an internet-accessed application based on CGI/Perl programming that designs PCR primers to search for deletion alleles in *Caenorhabditis elegans* gene knockout experiments and uses electronic PCR to search the entire genomic DNA sequence for potential false priming or multiple PCR amplification targets. Features such as the ability to target specific exons with the 'poison primer' approach and evaluation of primers with electronic PCR provide a flexible, web-based approach to design effective primers whilst minimizing the need for empirical optimization of PCR experiments.

Availability: Web access to this program is provided at <http://elegans.bcgsc.bc.ca/gko/aceprimer.shtml>.

As the list of genes targeted by the *C. elegans* gene knockout consortium grows (<http://elegans.bcgsc.bc.ca/knockout.shtml>), the time consuming process of designing and testing targeted PCR primers for deletion screening becomes a bottleneck in the gene knockout process. Our gene-knockout protocols require two nested sets of PCR primers to detect deletions affecting particular coding sequences. The PCR strategies used by consortium laboratories are summarized at <http://elegans.bcgsc.bc.ca/strategy.shtml>. Briefly, a relatively large genomic region (3–5 kb) containing the target gene is amplified with external primers, then re-amplified with more closely spaced nested primers. Deletions are detected by a reduction in the size of the secondary PCR product. The 'poison primer' technique is a refinement of this approach that uses multiplex PCR to aid the detection of smaller deletions that were often missed using standard screening methods (Edgley *et al.*, 2002). The poison primer technique involves the strategic placement of extra primers that poison the PCR by allowing the amplification of alternative products to compete with the full-length products and inhibit the secondary amplification. Deletions affecting the target region are detectable because they remove the poison primer binding sites and restore

the ability to amplify full-length PCR products. The poison primer technique enhances detection of small deletions in complex mixtures whilst targeting particular regions of the gene via placement of the poison primer binding sites.

The process of designing multiplex gene knockout primer sets is complex. An important consideration in a high-throughput environment is that PCR primer sets should work reproducibly well under standard conditions with minimal empirical optimization. Another challenge is that PCR primers may interact in unexpected ways. Use of multiple rounds of PCR and multiplex PCR are particularly sensitive to unwanted products produced by secondary priming elsewhere in the genome. We developed AcePrimer to automate the gene-based PCR primer design process and provide a means of presynthesis evaluation for potential unwanted PCR products. The annotation-directed primer design performed by AcePrimer requires targeting primers to encompass specific genomic features, in this case coding exons, whilst avoiding potential problem areas such as repetitive sequences, large introns and nested genes. AcePrimer was originally developed for use by the gene knockout consortium to design primer sets for all of the approximately 20 000 *C. elegans* genes predicted from the genomic DNA sequence. However, ongoing refinement of *C. elegans* gene predictions and sequence annotation and the continued need to design custom primer sets have led us to adapt AcePrimer for public use on individual genes.

AcePrimer uses the AcePerl PERL module (Stein and Thierry-Mieg, 1998) to extract gene-specific information from the current *C. elegans* AceDB database release and assemble a gene feature format (GFF) object that contains the functional annotation for specific genomic regions (<http://www.sanger.ac.uk/Software/formats/GFF/>). This information is then used to delimit the target DNA sequence to select primers that sequentially target each exon in the gene. Optimal primer sequences are predicted using PRIMER3 (Rozen and Skaletsky, 1998). The poison primer design is configured so each poison primer is situated within the target exon and oriented to pair with

*To whom correspondence should be addressed.

one of the external primers to produce an alternative PCR product suitable for out-competing the product of the two external primers.

A critical component of our primer design process, in particular with the multiplex poison PCR approach, is to evaluate whether the primers could potentially produce unexpected amplification products. The electronic PCR program (e-PCR) predicts PCR products by using a defined primer pair to scan a provided sequence file (Schuler, 1997). e-PCR was not designed for our application because it only does single pair-wise tests and does not predict products from multiplexed primers or from unexpected pairings, such as either the left or right primer alone priming a PCR product elsewhere in the genome. However, we were able to adapt e-PCR for use in AcePrimer by dividing the evaluation process into individual pair-wise comparisons. In this manner, all possible pair-wise primer combinations are tested sequentially against the *C. elegans* genomic DNA sequence. For example, a gene-knockout primer set for an experiment with two poison primers requires ten pair-wise e-PCR tests for the first-round amplification (four primers against the entire genome) and three steps for the second-round amplification (two primers against the external PCR amplicon). Based on the e-PCR test, primer sets predicted to produce unwanted PCR products smaller than 5 kb at any stage of the testing are rejected.

Numerous PCR primer design applications, including the core program PRIMER3, are available as source code or binary distributions. However, many are limited to particular operating systems, which may be a barrier to the casual user. A web interface offers the advantage of platform-independence and eliminates the need for local software installations. Several existing web sites, such as the PRIMER3 site (Rozen and Skaletsky, 1998), PrimerFinder (<http://eatworms.swmed.edu/~tim/primerfinder>) and Web Primer (<http://genome-www2.stanford.edu/cgi-bin/SGD/web-primer>) offer online primer design. These general purpose primer design utilities lack the ability to perform nested or multiplexed PCR primer design, automated retrieval of DNA sequence and annotation from the database and to perform a genome-wide evaluation of potential false priming. Using AcePrimer's annotation-directed design and e-PCR approaches, we were able to extend the functionality of the core program PRIMER3 so that all of these functions could be accomplished in a single step, with the required input limited to the name of the gene and, optionally, the primer design parameters. While AcePrimer is currently used for *C. elegans* primer design, future plans include adapting this approach to other species for which annotated genomic DNA sequence is available.

AcePrimer has enough user-configurable options to

be sufficiently flexible to accommodate the needs of *C. elegans* gene knockout experiments or to design gene-specific PCR primers for other applications. User configurable options include PRIMER3 design parameters such as oligonucleotide size, melting temperature, PCR product size and the optimal PRIMER3-assigned quality estimate, which is a measure of departure from various optimality criteria. Other options include the design and relative position of internal primers, whether to design one or two poison primers and whether to target a specific exon. The stringency of e-PCR is controlled by defining the number of sequence mismatches allowed for primer binding and the size of the perfect sequence match required at the critical 3' end of the primer.

The AcePrimer results are formatted as easily read and copied tabular output displayed in the browser window. Basic gene properties are reported along with a pointer to detailed information on the gene at <http://www.wormbase.org>, the online version of the *C. elegans* AceDB database. The primer design results list the primer size, location, melting temperature, PCR product size and calculations of the amount of coding DNA and the number of exons contained by the PCR product. The e-PCR tests are computationally intensive and time consuming. The user can select the relatively slow option of running e-PCR on all primers as they are created or to run e-PCR on selected primer sets after they are designed. The latter is accomplished by means of a standalone CGI application that is loaded by submitting an inline form provided as part of the AcePrimer output. Primer sets are evaluated in a separate window so the user may test multiple primer sets without having to re-submit the primer-design form.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Don Moerman, Mark Edgley and Erin Gilchrist for feedback on AcePrimer. We also acknowledge the British Columbia Cancer Agency, The Canadian Foundation for Innovation, the Canadian Institute for Health Research and Genome Canada.

REFERENCES

- Edgley, M., D'Souza, A., Moulder, G., McKay, S., Shen, B., Gilchrist, E., Moerman, D. and Robert Barstead, (2002) Improved detection of small deletions in complex pools of DNA. *Nucleic Acids Res.*, **30**, e52.
- Rozen, S. and Skaletsky, H.J. (1998) Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.
- Schuler, G.D. (1997) Sequence mapping by electronic PCR. *Genome Res.*, **7**, 541–550.
- Stein, L.D. and Thierry-Mieg, J. (1998) Scriptable access to the *Caenorhabditis elegans* genome sequence and other ACEDB databases. *Genome Res.*, **8**, 1308–1315.