BIOINFORMATICS APPLICATIONS NOTE Vol. 17 no. 1 2001
Pages 98–99

PrimeArray: genome-scale primer design for DNA-microarray construction

Günter Raddatz[1], Michaela Dehio[1, 3], Thomas F. Meyer[1, 2] and Christoph Dehio[1, 3]

[1]Max-Planck-Institut für Biologie, Abteilung Infektionsbiologie, Spemannstraße 34, 72076 Tübingen, Germany, [2]Max-Planck-Institut für Infektionsbiologie, Abteilung Molekulare Biologie, Monbijoustr. 2, 10117 Berlin, Germany and [3]Biozentrum der Universität Basel, Abteilung Molekulare Mikrobiologie, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Received on April 13, 2000; revised on June 23, 2000; accepted on July 25, 2000

ABSTRACT

Summary: PrimeArray is a Windows program that computes oligonuceotide primer pairs for genome-scale gene amplification by the Polymerase Chain Reaction (PCR). The program supports the automated extraction of coding sequences (CDS) from various input-file formats and allows highly automated primer pair-optimization.

Availability: The program is freely available for non-profit use via request from the authors.

Contact: christoph.dehio@unibas.ch

DNA microarrays enable a parallel analysis of thousands of genes providing insights into gene function (via expression monitoring) and the relevance of genetic loci for phenotypic traits. For the production of microarrays, DNA can either be synthesized on a solid support by spatially patterned, light-directed combinatorial chemistry (Lipshutz *et al.*, 1999), or can be deposited in a pre-synthesized form (i.e. as PCR-products) onto a surface by various spotting techniques (Cheung *et al.*, 1999). Spotted DNA microarrays were shown to give rise to reliable and highly reproducible results and facilities for their production are implemented in an increasing number of academic and industrial setting (Cheung *et al.*, 1999). PCR-based gene amplification is typically a crucial step in the production of spotted DNA microarrays. Several software tools have been written to design and optimize primers for the particular purpose of large-scale sequencing projects, i.e. PRIDE (Haas *et al.*, 1998), PRIMER MASTER (Proutski and Holmes, 1996) and PRIMO (Li *et al.*, 1997), which all employ sophisticated algorithms to avoid unfavorable primer–primer or primer–template interactions. However, while the number of completed genome sequences is steadily growing, none of these programs is specialized on the automated design of primer pairs for genome-scale gene amplification as desired for the production of DNA microarrays.

Here we present the program PrimeArray specifically designed for this purpose. PrimeArray is written in Delphi5.0 and runs under Windows 95/98/NT4.0. The program has a comfortable graphical user interface and allows all parameters of the computations to be configured by the user (Figure 1). Coding sequences can be extracted from various input file formats (i.e. GenBank-format) without time-consuming cut-and-paste pre-processing of the downloaded files containing annotated DNA sequence information. Computation of primer pairs is done by iteratively shifting primer candidates along the sequence until melting temperature, GC-content and absence of unfavorable self- and primer–primer interactions are fulfilled. Constructed pairs of forward and reverse primers and the predicted PCR product size can be written to an Excel-spreadsheet as well as to an ASCII-file. These features of PrimeArray combined with a minimal user interaction make this program particularly useful for the large scale design of PCR-primer pairs for the purpose of DNA-microarray production.

Melting temperature of the primer–template interaction is computed by applying the rules of Suggs (Suggs *et al.*, 1981), which assign 2 °C for an A/T base pair and 4 °C for a G/C base pair. Alternatively, the program allows the application of a refined nearest-neighbor method based on the parameters of Sugimoto *et al.* (1996). Interaction between forward and reverse primer as well as self-complementarity of primers like stem-loops are determined by computing the highest number of neighboring complementary bases. Also, checks for GC-content and occurrence of secondary binding sites are performed. Further options provided by PrimeArray include analysis of user defined primer-pairs, user-controlled optimization of individual primers with distinct optimization parameters and definition of a maximum region of interest of the underlying nucleotide sequence. Optimization of primer pairs is done by applying the following protocol: first, the starting base of the forward-primer is shifted along the

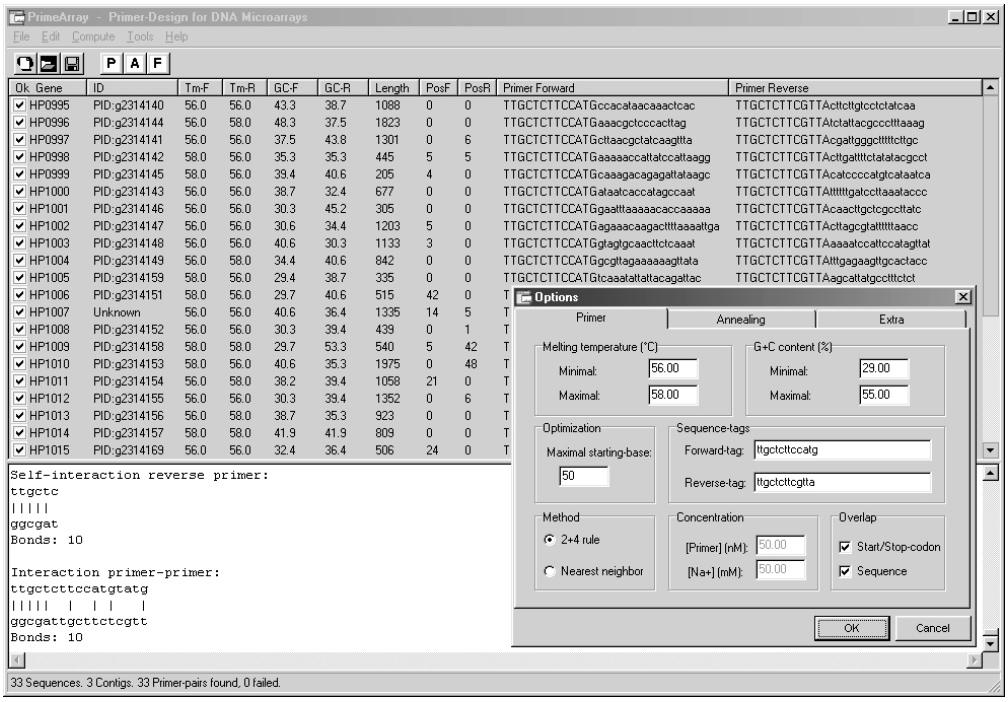

Fig. 1. Screenshot of the graphical user-interface of PrimeArray. The GUI is split into the results window (grey) showing designed primers and the messages window (white) which shows details of unfavorable interactions.

sequence of the CDS, until all criteria for a valid single primer are fulfilled. Then the starting base of the reverse primer is shifted from the end of the CDS until the criteria for a single valid primer and also a sufficient low interaction between a forward and the respective reverse primer are fulfilled. If no acceptable reverse primer can be found, the forward primer is shifted further, until the criteria for a single primer are fulfilled again and the procedure for the reverse primer is repeated. This strategy is repeated iteratively. If the length of amplified DNA-fragments should be limited, a cut-off can be defined. PrimeArray also allows the user to freely define tag sequences, which are appended to the primer candidates.

Reliability of the designed gene-specific primer pairs was analyzed on a subset of 475 genes of the genome (1490 genes) of the pathogenic bacterium *Helicobacter pylori*. In a first round of primer optimization PrimeArray designed primer pairs for 467 genes giving rise to 459 successful PCR reactions (validated by the production of a single PCR product matching the calculated size). For the remaining 8 genes, for which the program could not automatically design primer pairs due to the stringency of chosen parameters, those parameters were stepwise loosened until primer pairs could be computed. 7 out of the 8 resulting primer pairs gave rise to the expected PCR products. Summarizing these results, we could demonstrate a success-rate of >98% (466 out of 475) for correct PCR reactions by primer pairs designed by PrimeArray, suggesting a highly reliable performance of this software tool. PrimeArray should thus aid to a cost-effective design and production of spotted DNA microarrays.

REFERENCES

Cheung,V.G., Morley,M., Aguilar,F., Massimi,A., Kucherlapati,R. and Childs,G. (1999) Making and reading microarrays. *Nature Genet.*, **21** (Suppl.), 15–19.

Haas,S., Vingron,M., Poustka,A. and Wiemann,S. (1998) Primer design for large scale sequencing. *Nucleic Acids Res.*, **26**, 3006–3012.

Li,P., Kupfer,K.C., Davies,C.J., Burbee,D., Evans,G.A. and Garner,H.R. (1997) PRIMO: a primer design program that applies base quality statistics for automated large-scale DNA sequencing. *Genomics*, **40**, 476–485.

Lipshutz,R.J., Fodor,S.P., Gingeras,T.R. and Lockhart,D.J. (1999) High density synthetic oligonucleotide arrays. *Nature Genet.*, **21** (Suppl.), 20–24.

Proutski,V. and Holmes,E.C. (1996) Primer Master: a new program for the design and analysis of PCR primers. *Comput. Appl. Biosci.*, **12**, 253–255.

Suggs,S.V., Hirose,T., Miyake,E.H., Kawashima,M.J., Johnson,K.I. and Wallace,R.B. (1981) Using purified Genes. In Brown,D.D. (ed.), *ICN-UCLA Symposium on Developmental Biology* vol. 23, Academic Press, New York, 683.

Sugimoto,N., Nakano,S., Yoneyama,M. and Honda,K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res.*, **24**, 4501–4505.