BIOINFORMATICS APPLICATIONS NOTE Vol. 17 no. 7 2001
Pages 656–657

MEDUSA: large scale automatic selection and visual assessment of PCR primer pairs

Raf M. Podowski and Erik L. L. Sonnhammer

Center for Genomics Research, Karolinska Institutet, S-171 77 Stockholm, Sweden

Received on January 12, 2001; revised on March 14, 2001; accepted on March 16, 2001

ABSTRACT

Summary: MEDUSA is a tool for automatic selection and visual assessment of PCR primer pairs, developed to assist large scale gene expression analysis projects. The system allows specification of constraints of the location and distances between the primers in a pair. For instance, primers in coding, non-coding, exon/intron-spanning regions might be selected. Medusa applies these constraints as a filter to primers predicted by three external programs, and displays the resulting primer pairs graphically in the Blixem (Sonnhammer and Durbin, *Comput. Appl. Biosci.* **10**, 301–307, 1994; http://www.cgr.ki.se/cgr/groups/sonnhammer/Blixem.html) viewer.

Availability: The MEDUSA web server is available at http://www.cgr.ki.se/cgr/MEDUSA. The source code and user information are available at ftp://ftp.cgr.ki.se/pub/prog/medusa.

Contact: Erik.Sonnhammer@cgr.ki.se

The ongoing effort of functional genomics strives to unravel the function of all genes in a genome. A first step of functional characterization of a gene is to identify its expression pattern. Most approaches for measuring expression are based on PCR products (Freeman, 2000), and thus require PCR primers that are highly specific for the mRNA of the gene studied and give high yields of product. A common problem are products from genomic DNA, which can be avoided by positioning the primers on opposite sides of a known intron and sizing the product, or by placing the primer over a splice site. Carrying out this type of selection manually is a tedious and error-prone task. A number of PCR primer design programs exist (Hillier and Green, 1991; http://www.genome.wi.mit.edu/genome_software/other/primer3.html; Wisconsin Package Version 10, http://www.gcg.com/; EMBOSS Software Package, http://www.uk.embnet.org/Software/EMBOSS/). However, none of these take gene structure or distance constraints into account. Other, related programs exist for designing DNA sequencing primers (Hillier and Green, 1991; Li *et al.*, 1997).

In order to automate efficient and reliable PCR primer selection using gene structure or distance constraints, MEDUSA, a primer pair parsing system was created. MEDUSA accepts a sequence template, feature information (exons, etc.), PCR primer design criteria and results display requests. It obtains PCR primer pairs from a number of oligo design programs, filters them based on user-defined criteria and combines the results to produce a comprehensive, informative view of the selections. The user can specify a wide variety of constraints on allowed primer locations relative to the specified Coding Sequence (CDS), Untranslated Terminal Regions (UTR), and splice site positions using a simple format.

MEDUSA is available in two formats. A web server is useful for small-scale selection. For large-scale projects, the platform-independent python (http://www.python.org) script and user documentation can be downloaded from an ftp server for local usage, which requires the user to have a working version of the python interpreter, at least one of the primer selection programs used by MEDUSA, and Blixem (Sonnhammer and Durbin, 1994; http://www.cgr.ki.se/cgr/groups/sonnhammer/Blixem.html) for visual display.

Currently, MEDUSA obtains primer pairs from a desired combination of three programs: primer3, GCG prime and EMBOSS prima. These programs are executed by MEDUSA, and their output is parsed directly; additional primer design programs can easily be added. MEDUSA user input is divided into a number of categories, described in detail in a README file available on the www and ftp servers. The only required input is the sequence in plain, single or multi-line format following a sequence start tag line.

The second input category includes selection flags for execution of specific primer design programs and masking of the sequence with repeatblaster or repeatmasker programs, allowing low complexity regions to be excluded from the primer site selection, as well as template sequence features. These consist of the start and stop positions of the coding region (or region of special interest) and the number and position of any exons or secondary regions of interest.

The third input category consists of PCR primer and

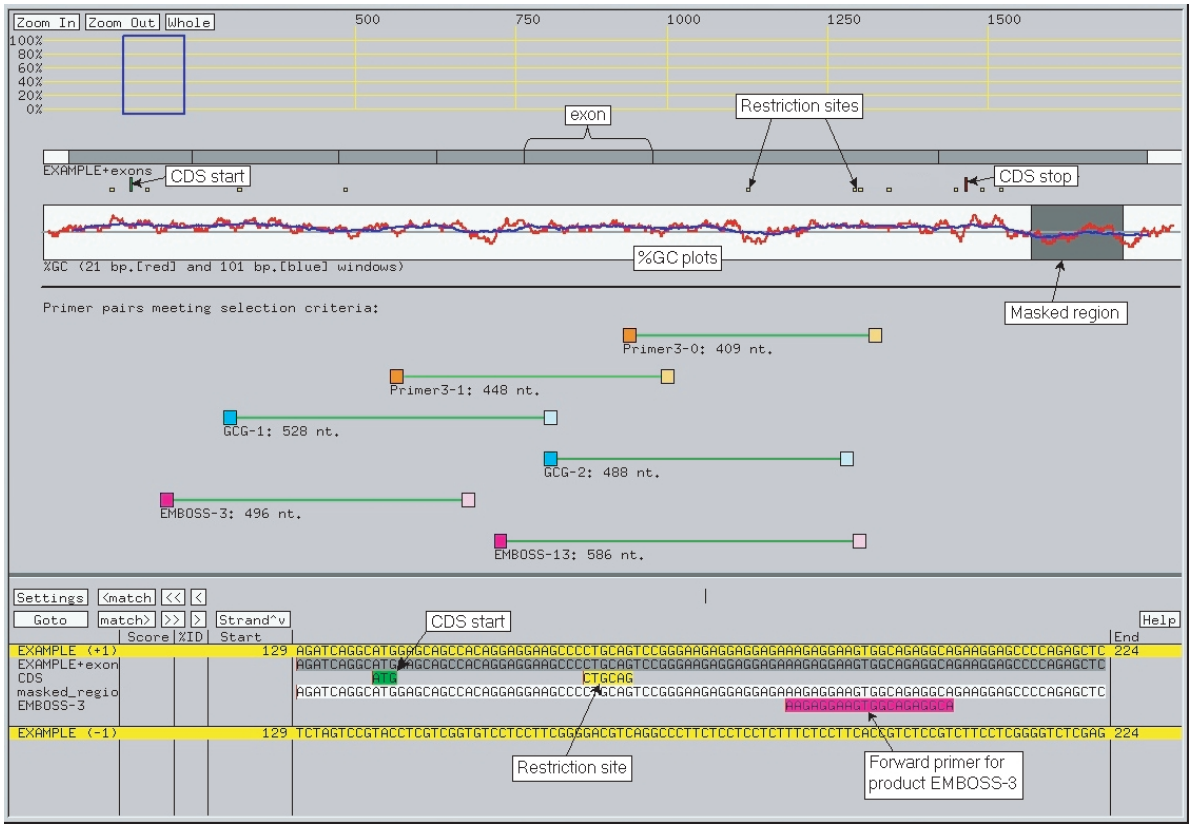

Fig. 1. MEDUSA selected primer pairs viewed in Blixem. Segment features are displayed both graphically in the top panel and as a sequence in the bottom panel. The first bar from the top indicates the exon structure. CDS start/stop and restriction sites are marked below. Next, profiles of %GC content for 21 and 101 bp windows. Segments masked as repeat regions are shown as dark regions under the %GC profiles. Below this, the selected primer pair products are shown together with length information.

product selection parameters. The recommended input includes primer and product size ranges and a primer melting temperature range. Additionally, input parameters specific to any of the three primer selection programs can be included in the input. The user can provide a list of restriction enzymes and recognition sites in order to screen the template and product for specific cleavage sites. These sites are listed in the results and shown in the graphical output as additional information.

Finally, MEDUSA combines, filters and sorts products based on the user specified criteria for the PCR product's position with respect to the CDS region and exon splice sites location. The selected primer pairs can then be viewed in Blixem (see Figure 1), along with template sequence features. The data is passed from MEDUSA to Blixem in a simple general-purpose format called SFS (E.L.L.Sonnhammer, unpublished observations). The user can examine the primer and product sequences against the template's features for a more intelligent selection or to help resolve PCR problems. The combination of the automatic primer filtering program, MEDUSA, and the graphical overview provided by Blixem makes selection of PCR primer pairs efficient and reliable.

REFERENCES

Freeman,T. (2000) High throughput gene expression screening: its emerging role in drug discovery. *Med. Res. Rev.*, **20**, 197–202.

Hillier,L. and Green,P. (1991) OSP: a computer program for choosing PCR and DNA sequencing primers. *PCR Methods Appl.*, **1**, 124–128.

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.

Sonnhammer,E.L.L. and Durbin,R. (1994) A workbench for large-scale sequence homology analysis. *Comput. Appl. Biosci.*, **10**, 301–307.