PIRA PCR designer for restriction analysis of single nucleotide polymorphisms

Xiayi Ke, Andrew Collins and Shu Ye

Division of Human Genetics, University of Southampton, Duthie Building, Southampton General Hospital, Southampton SO16 6YD, UK

ABSTRACT

Summary: Primer-introduced restriction analysis (PIRA-PCR) is widely used to detect Single Nucleotide Polymorphisms (SNPs). To create artificial Restriction Fragment Length Polymorphism (RFLP), a mismatch is usually introduced near the $3'$ end of the primer that is close to the mutation of interest. We describe in this report a www-based computer program that screens for the suitable mismatches, designs the primers, lists the appropriate restriction enzymes and other related information.

Availability: The computer program, with related descriptions, is available at http://cedar.genetics.soton.ac.uk/public_html/primer2.html.

Contact: xke@soton.ac.uk

PIRA-PCR is one of the most widely used techniques in SNP detection (Haliassos *et al.*, 1989; Jacobson and Moskovits, 1991). The method introduces an artificial restriction site into a PCR product by the use of a primer with a single-base mismatch close to its $3'$ end. No www-based computer tools have been reported that provide users with convenient access and use although a computer program, written in BASIC, for designing the mismatches is available (Davidow, 1992). Furthermore, the above program was written mainly from the perspective of designing mismatches that provide artificial RFLPs. In this communication, we describe a www-based computer tool that firstly selects all the mismatches and then uses a set of criteria, which can be adjusted by users, to design primers based on the selected mismatches.

The program searches all mismatches on both sides of the mutation and on both strands that would create an artificial restriction site either for the wild-type or the mutant but not for both. The recognition sequence must span from a mismatch to the mutant/normal nucleotide and can also stretch outwards in both directions. All Type II enzymes in the REBASE Restriction Enzyme Database (http://rebase.neb.com) are targeted. Once a list of mismatches and their associated restriction enzymes are obtained, the mismatches and related information, such as whether a mismatch is found with the wild-

type or the mutant DNA and whether it is on the sense strand or anti-sense strand, are passed to the modules that design the two necessary PCR primers. Primer melting temperature is calculated using the nearest neighbor parameters (Breslauer *et al.*, 1986) and the formula given by Rychlik *et al.* (1990). The algorithm for calculating complementarity is similar as described by Rozen and Skaletsky (2000). The program looks for the primer pairs that meet the user requirements in the following way: (1) select all primer pairs that meet the basic user requirements (primer size range, PCR product size range, Guanine and Cytosine (GC) content range, melting temperature (T_m) range and maximum complementarity/$3'$ complementarity), (2) select the best primer pairs based on whether they meet the optimum conditions as set by the user (optimum primer size, optimum PCR product size, optimum T_m), (3) among the best primer pairs, those with the same T_m are output first. Java was used for the implementation.

To access the service, users are expected to use newer versions of either Netscape Navigator (version 3.0 or later) or Microsoft Internet Explorer (version 4.0 or later). The users are asked to input a source sequence (up to 1000 bases), indicate the location of the mutation in the sequence and the normal and mutant nucleotides. Various primer-design criteria can be set by users otherwise the default values will be used. The source sequence can be a local sequence segment or a GenBank/EMBL DNA sequence. 'A', 'T', 'G' and 'C' are accepted as they are. All ambiguity letters ('R', 'Y', 'M', 'K', 'S', 'W', 'H', 'B', 'V', 'D' and 'N') will be expanded during the search. Symbols '∗' and '−' are treated as 'N'. Case is not significant and all other characters, including spaces and digits, are ignored.

Figure 1 is a sample output of using the program to search mismatched PCR primers and restriction enzymes for the G-174C polymorphism in the *Homo sapiens* interleukin-6 (IL-6) gene. Using the AF048692 GenBank sequence and the default settings on the web interface, a total of seven entries (all with a mismatch introduced except the first one) have been found. For each recognition

```
*******************************************************************
Mismatch T introduced at position of 475
PCR primers recommended
5'          Primer pair sequence                      3'      Tm
451         CACTTTTCCCCCTAGTTGTGTGTCTTTC               476     64
650         GAGCCTCAGACATCTCCAGTCCTATATT               623     64
Recognition sequence: TCGA
Locations: 475, 660, 772
Enzymes: CVISIII, ESABC3I, HPYV, HPY26II, HPYF14III, HPYF16I, HPYF23I, HPYF24I, HPYF26III,
HPYF30I, HPYF35I, HPYF40II, HPYF42IV, HPYF45I, HPYF49I, HPYF52I, HPYF59III, HPYF62II,
HPYF64I, HPYF65II, HPYF66IV, HPYF71I, HPYF73II, HPYJP26II, PPAAII, TAQ20I, TBR51I, THA3I,
THTOK4A2I, THTOK6A1I, TFLI, TSC4AI, TSP32I, TSP32II, TSP358I, TSP505I, TSP510I,
TSPAK13D21I, TSPAK16D24I, TSPNI, TSPVI4AI, TSPVIL3I, TTH24I, TTHHB8I, TAQI, TTHRQI
*******************************************************************
Mismatch C introduced at position of 474
PCR primers recommended
5'          Primer pair sequence                      3'      Tm
451         CACTTTTCCCCCTAGTTGTGTGTCTCGC               476     68
655         AGAATGAGCCTCAGACATCTCCAGTCCT               628     68
Recognition sequence: TCGCGA
Locations: 473
Enzymes: AMAI, BSP68I, MLU9273I, MLUB2I, SALDI, SBO13I, SNA3286I, SPOI, UBA1117I,
UBA1386I, NRUI, VCHO70I

PCR primers recommended
5'          Primer pair sequence                      3'      Tm
451         CACTTTTCCCCCTAGTTGTGTGTCTCGC               476     68
655         AGAATGAGCCTCAGACATCTCCAGTCCT               628     68
Recognition sequence: CGCG
Locations: 474
Enzymes: ACCII, BCE3129I, BCEBI, BCERI, BEPI, BPU95I, BSH1236I, BSP50I, BSP70I, BSP123I,
BSPJ76I, BSTFNI, BSTUI, BSU1192II, BSU1193I, BSU1532I, BSU6633I, BSUEII, BTKI, CPA1150I,
CPAAI, CSP68KVI, CSPKVI, FAUBII, FSPMI, HIN1056I, HPYF14I, HPYF15I, HPYF26I, HPYF52II,
HPYF64IV, MVAAI, MVNI, PFLAI, SCEAI, SELI, THAI, TMAI, UBA1321I, UBA1404I, UBA1405I,
FNUDII,UBA1446I
*******************************************************************
```

Fig. 1. A representative output of the program. With the first entry, line 1 indicates whether and where the mismatch is introduced; lines 2–5 describe the recommended PCR primers; line 6 is the restriction enzyme recognition sequence associated with the concerned mismatch; line 7 gives all the locations of the recognition sequences throughout the source sequence; line 8 to end lists all relevant Type II restriction enzymes. The second entry shows a mismatch produces two different recognition sequences, each of which is presented with associated information about its primer, locations and relevant enzymes.

sequence associated with the introduced mismatch, one primer pair is output. This in normal situations would suffice. If more than one primer pair is needed, the setting ('Number of primer pairs') on the web interface can be increased accordingly. The output also gives all the locations of the recognition sequence in the source DNA.

The information helps the user decide whether or not to choose a particular enzyme. If more than one recognition sequence for a particular restriction enzyme is identified in the PCR product of the two recommended primers, users can usually exclude the undesired ones from the PCR product by decreasing the product size settings on the web interface. The default setting for the number of bases allowed for mismatches on both sides of the mutation is 4. This setting is capable of searching the recognition sequences of most of the Type II enzymes. If enzymes with longer recognition sequence, such as AloI and HaeIV, are required, the value of the setting ('Number of bases for mismatches on both sides of SNP') can be increased. The upper bound for this setting is 14 since a value larger than 14 is both unnecessary and consumes a lot of server resources.

REFERENCES

Breslauer,K.J., Frank,R., *et al.* (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA*, **83**, 3746–3750.

Davidow,L.S. (1992) Selecting PCR designed mismatch primers to create diagnostic restriction sites. *Comput. Appl. Biosci.*, **8**, 193–194.

Haliassos,A., Chomel,J.C., *et al.* (1989) Detection of minority point mutations by modified PCR technique: A new approach for a sensitive diagnosis of tumor-progression markers. *Nucleic Acids Res.*, **17**, 8093–8099.

Jacobson,D.R. and Moskovits,T. (1991) Rapid, nonradioactive screening for activating ras oncogene mutations using PCR-primer introduced restriction analysis (PCR-PIRA). *PCR Meth. Appl.*, **1**, 146–148.

Rozen,S. and Skaletsky,H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. *Meth. Mol. Biol.*, **132**, 365–86.

Rychlik,W., Spencer,W.J., *et al.* (1990) Optimization of the annealing temperature for DNA amplification *in vitro*. *Nucleic Acids Res.*, **18**, 6409–6412.