Nucleic Acids Research, 1994, Vol. 22, No. 15 3247–3248

Gene walking by unpredictably primed PCR

Orlando Domínguez[1,*] and Carlos López-Larrea[1,2]
[1]Servicio de Inmunología, HCA, E-33006 Oviedo and [2]Departamento de Inmunología, UCM, Madrid, Spain

Received June 3, 1994; Revised and Accepted July 1, 1994

Several methods have been described for gene walking by PCR on uncloned DNA (2–13). In general, they suffer from limitations: they either avoid high-complexity whole genomes, require delicate manipulation or give contaminated products. This report describes a reliable and simple PCR (1) adaptation for gene walks from any specific genomic locus: unpredictably primed PCR (UP-PCR). Besides its similarity to some of those predecessors, UP-PCR offers several advantages: quality of products, simplicity, general applicability and a methodical approach to optimization. It employs undigested mammalian genomic DNA as the original template source and defined walking primers of known working range; it yields remarkably clean results, and can produce a gene walk in a single working day.

It employs two successive PCRs, for which four oligos are needed: two sequence specific primers (SPs) and two universal walking primers (WPs) (Figure 1 part 1). The SPs are complementary to the known sequence from which one intends to walk. Both have the same orientation: 3′ end towards the unknown region. They can be called outer specific primer (SP) and inner specific primer (iSP), respectively (Figure 1 part 1). The latter gives the advantages of a nested primer and it is used at the second reaction. In order to save limited sequence information it has been currently made shorter than the SP (Table 1).

The WPs are oligonucleotides of defined and artificial sequence. Based on preliminary experimentation they have been given a similar basic design (T_mG_n) in order to fulfil the following requirements: promiscuous character for the first hybridization step (see below) and selectivity for subsequent PCR annealings. WPs differ from one another in their melting temperature (*Tm*), having therefore distinct working ranges, and at small changes at their 3′ ends (Table 1). A ~30-mer and a 17-mer are used at the first and nested reactions, respectively (see below). The sequence of the short WP (sWP) is contained within that of the larger WP (Table 1, Figure 1 part 1).

Reactions were performed after preparing the following mixes: Sample mix, made in duplicate: 20 ng human or murine whole undigested genomic DNA (gDNA), 20 pmol WP, dH_2O to 10 μl and mineral oil. For a control with pBluescript plasmid (Stratagene), its genome was present at 100 molecules per microliter of final reaction, among 20 ng of human gDNA as background to mimic mammalian single copy locus representation.

Polymerase mix (volumes per individual sample): 4 μl 5× buffer (42.5 mM $(NH_4)_2SO_4$, 62.5 mM Tris–HCl pH 8.3, deionized Tween-20 and Triton X-100 (each at 0.25%, sterilized by filtration through 0.2 μm and UV irradiation for 10 minutes), 2 μl dNTP mix (with 2 mM each dNTP) and 0.08 μl (0.4 U) Taq DNA pol (Amplitaq, Perkin-Elmer).

Both replicas of sample mix were incubated for 30–60s at 90°C in a thermocycler (GeneAtaq, Pharmacia-LKB) and suspended at 80°C. The next components were then added: 2 μl of $MgSO_4$, from a 15 mM solution to the first tube and from a 30 mM stock to the second (solutions sterilized as already described), and 6 μl of polymerase mix. Tubes were inmediately chilled in a water bath (15–25°C) and incubated at room temperature for 10 minutes. This low stringency annealing and

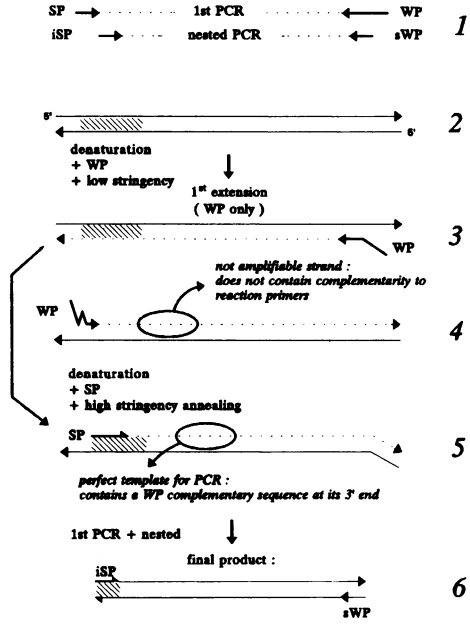

Figure 1. Schematic representation of the UP-PCR. 1. Overview with relative position of primers. 2. original genomic double stranded template, showning the known sequence from the desired locus as a crosshatched area. The template is denatured and a subsequent low-stringency annealing step allows the ubiquitous and unpredictably extension 3 and 4) of the only primer present: the WP. Dotted lines represent strands synthesized at that step. 3, WP extension that generates a sequence complementary to the SP. It will be selected at the following cycle. 4, WP extension that will not be amplified. 5, Extension of the just introduced SP generating a sequence complementary to the WP. By first time there is such an artificial sequence in the sample. PCR starts. 6, Specific walk obtained after the second or nested reaction.

* To whom correspondence should be addressed

Table 1. Primers used in this work

Primer [SP or WP]	Locus	Sequence (5'→3')	L[1]	Tm[2]	WP[3]	Ta[4]
C69 [WP]		$T_{11}GT_2GT_2GTG_2TT$	33	67.6		
D74 [WP]		$T_{11}GT_2GT_2GTG_4TGT$	28	62.3		
C47 [sWP]		$T_2GT_2GT_2GTG_3$	17	41.2		
K81as(255)[5] [SP]	human K-ras2	ATTTATGGCAAATACACAAAGAAAGC	26	54.4	D74	63
K81as(222)[5] [iSP]		AGTCCTCATGTACTGGTCCC	19	47.5	C47	55
mH4.4as(54)[5] [SP]	mouse Complement factor H	TGCTGCACAAACAGTCCATAATATAAGC	28	58.8	D74	62
mH4.4as(26)[5] [iSP]		CAAATAATTCTTGCTGACA	19	40.6	C47	54
M13(-72)[5] [SP]	lacZ in pBluescript	GCTGCAAGGCGATTAAGTTGGGTA	24	60.3	C69	66
M13(-26)[5] [iSP]		GACGTTGTAAAACGACG	17	40.9	C47	54

(1) Length. (2) *Oligo Tm* (3) Walking primer used on the respective reaction. (4) Annealing temperature of the respective PCR. (5) Position of the 5′ end relative to the initiation codon on cDNA. (6) Position of the 5′ end downstream to the *Hin*dIII site of M13mp18.

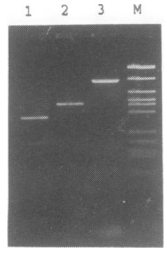

Figure 2. Typical UP-PCR products. 1, human K-ras2, walk from exon 2 into the previous intron, 318 bp. 2, murine complement factor H, walk from its leader sequence into the promoter region, 453 bp. 3, walk inside the lacZ of pBluescript, across the MCS, 853 bp. Fragment sequences are in agreement with EMBL database accession numbers L00046, J02891 and X52328, respectively. 4, PhiX174 RF *Rsa*I digest as size marker.

extension step will be referred to as first extension (Figure 1, parts 3 and 4) and it was considered complete after a final 30 s incubation at 75°C on the thermocycler. Samples were held at 90°C and 2 μl of 10 μM SP (20 pmol) were finally added. The first PCR started with the following program: 30 s at 94°C, 30 s at *Ta* (annealing temperature; see below) and 30 s at 75°C, for 30 cycles followed by a final extension of 7 minutes at 75°C. Some faster machines may require longer segment times. The Mg^{2+} gradient between tubes helped optimization.

The nested PCR used 1:1000 dilutions from the 1st reaction as 5× concentrated template, sWP and iSP as primers (each at 0.5 μM), and 25 to 30 cycles in otherwise standard conditions (same buffer and 1.5 mM Mg^{2+}). Its *Ta* has also to be stringent and is defined below.

The *Ta* is the most critical parameter for UP-PCR. Whereas a standard PCR can yield specific products within a *Ta* range of 10−15°, UP-PCR requires a much lower variation (~2°C). Under the chosen conditions it was found that a good correlation between the reaction *Ta* and the theoretical *Tm* calculation is given by the program *Oligo* (v 3.4, MedProbe), which is based on nearest-neighbour ΔG values (14). For convenience the *Tm* under default parameters was obtained uncorrected and taken as a reference value that is lower than the *Ta*. This program was sometimes used to decide primer length, as a function of *Tm*. Other programs using the same algorithm should give similar values. An alternative that gives equivalent results, only for the large SPs, is this adaptation of a published equation (15):

$$Tm = 62+24[(nG+nC)/l]-(360/l)$$

The next rule of thumb was empirically obtained and predicts the first reaction *Ta*

$$Ta = Tm^{SP}+4 \sim Tm^{WP},$$

where *Tm* values refer to *Oligo* estimations. Therefore we choose a WP with an *Oligo Tm* about 4° higher than the corresponding *Tm* of the given SP. For the nested PCR both primer *Tm* are balanced and *Ta* = *Tm*+14° has been used.

Regarding the first PCR it is advisable to stick to fixed conditions, leaving the *Ta* as the only factor to optimize. Absence of products, specific or not, can mean either low or high *Ta*. From the initial value given by the rule of thumb, *Ta* can be varied in three degree steps until specific products are found. In this manner any new walk is normally obtained in no more than three trial reactions. In all cases where the initial *Ta* failed to yield specific fragments, increasing it proved to be the correct choice, sometimes in combination with a higher *Tm* WP while keeping the same SP, suggesting that the predictions are occasionally underestimates.

The reliability of UP-PCR for gene walking extends to every locus tested to date: different unpublished human and rodent sequences are among the shown examples (Figure 2). It has been used to access both promoter and intronic regions and consecutive walking steps were taken occasionally as desired. In every case the fragments proved to be correct after sequencing.

The WPs shown (Table 1) have been successfully used at every different walk tested within their working *Ta* range. In fact the WP needs to fulfil a loose criterion of complementarity to produce results in the first extension: three terminal 3′ bases are the only fixed condition, apart from variable base-pairing stabilization from the rest of the primer (not shown). D74 has given clean products from 61 to 63°C, inclusive, whereas C69 worked from 63 to at least 66° (not shown).

The unpredictable feature applies to the position from which the PCR primer of arbitrary sequence extends distally to the locus specific oligonucleotide. The high stringency needed limits the product size range to about 500 bp (Figure 2). Longer fragments not visible with ethidium bromide staining have not been explored.

ACKNOWLEDGEMENTS

O.D. and this project have been supported by FISBII and CAICYT 92/PB-1042 grants, respectively.

REFERENCES

1. Saiki,R.K., (1985) *Science* **230**, 1350−1354.
2. Ochman,H., *et al.* (1988) *Genetics* **120**, 621−623.
3. Triglia,T., *et al.* (1988) *Nucleic Acids Res.* **16**, 8186.
4. Roux,K.H. and Dhanarajan,P. (1990) *Biotechniques* **8**, 48−57.
5. Rosenthal,A. and Jones,D.S.C. (1990) *Nucleic Acids Res.* **18**, 3095−3096.
6. Arnold,C. and Hodgson,I.J. (1991) *PCR Meth. Appl.* **1**, 39−42.
7. Copley,C.G., *et al.* (1991) *Biotechnology* **9**, 74−79.
8. Bloomquist,B.T., *et al.* (1992) *DNA Cell Biol.* **11**, 791−797.
9. Parker,J.D., *et al.* (1991) *Nucleic Acids Res.* **19**, 3055−3060.
10. Parks,C.L., *et al.* (1991) *Nucleic Acids Res.* **19**, 7155−7160.
11. Kaufmann,J. and Klein,A. (1992) *Nucleic Acids Res.* **20**, 4445−4450.
12. Shyamala,V. and Ames,G.F. (1993) *Methods Enzymol.* **217**, 436−446.
13. Jones,D.H. and Winistorfer,S.C. (1993) *PCR Meth. Appl.* **2**, 197−203.
14. Rychlik,W., *et al.* (1990) *Nucleic Acids Res.* **18**, 6409−6412.
15. Baldino,F., *et al.* (1989) *Methods Enzymol.* **168**, 761−777.