AMPLIFICATION OF CDNA ENDS USING PCR SUPPRESSION EFFECT AND STEP-OUT PCR.

Mikhail Matz (1*), Naila Alieva (1), Alex Chenchik (2), Sergey Lukyanov (3)

(1) Whitney Laboratory, University of Florida, 9505 Ocean Shore Blvd, St Augustine, FL 32080-8610, USA

(2) BD Biosciences Clontech, 1020 East Meadow Cir, Palo Alto CA 94303-4230, USA

(3) Evrogen GSC, Shemiakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, Miklukho-Maklaya 16/10, 117997 Moscow, Russia

** corresponding author, Email: matz@whitney.ufl.edu*

1. Introduction

Procedures outlined in this chapter require the cDNA prepared according to one of the two methods described in the file "Amplification of representative cDNA samples….". We will be referring to these two methods of cDNA synthesis and amplification as "**method A**" (the technique based on classical double-stranded cDNA synthesis and subsequent adapter ligation, implemented in Marathon kit by Clontech) and "**method B**" (based on template-switching effect, implemented in SMART cDNA synthesis kit by Clontech). These methods not only allow for amplification of representative cDNA populations from microscopic tissue samples, but also provide an excellent starting point for amplification of unknown flanks of a known cDNA fragment. There are many techniques designed for this purpose **(*1-10*)**, which among gene hunters go under the generic name RACE (Rapid Amplification of cDNA Ends). It must be remembered, though, that this name formally belongs to a particular technique introduced by Frohman et al **(*8*)**.

The method of RACE described here is based on three key principles. First, to achieve required specificity (and therefore sensitivity), we employ two-step PCR using nested gene-specific primers. The necessity of this is usually under-estimated in RACE, since historically the heaviest non-specific amplification in RACE was stemming from ubiquitous adapter-specific primers rather than gene-specific ones. Our own techniques were developed for single-step RACE **(*10, 4*)**, but they are only as good as the specificity of the single gene-specific primer that they rely upon. At the current stage of methodology development, it is non-specific amplification originating from gene-specific primer that often limits the power of the RACE, so we presently switched back to using the good old two-step nested PCR to get rid of this problem.

Second, the design of adapters and adapter-specific primers evokes so-called PCR-suppression effect, or PS-effect, to prohibit amplification of molecules that do not contain annealing site for a gene-specific primer **(*11*)**. The PCR-suppression technology is based on the observation that molecules flanked by inverted terminal repeats at least 40 bases long are amplified very inefficiently when the single primer used for amplification corresponds to the distal half of the repeat **(*10, 12*)**, or when a long primer

corresponding to the whole repeat sequence is used in low concentration **(*13*)**. The reason for this is the equilibrium between productive PCR primer annealing and non-productive self-annealing of the fragment's complementary ends, which arises within each PCR cycle at the primer annealing stage and in the above two cases is markedly shifted towards self-annealing **(*11*)**.

The third principle is step-out PCR **(*4*)**. In our protocol, this trick is applied during the second stage of amplification. It consists in substituting the adapter-specific primer used during the first stage for another long primer of which only the 3'-half corresponds to the adapter, specifically, to its distal half. In the resulting product, the non-matching 5'-portion of the new primer sequence becomes added to the distal end of the original adapter. This situation can be viewed as shifting of the primer annealing site to the outside of the original amplicon, hence the name "step-out". Although this procedure seems to be the contrary of nested PCR, is in fact analogous to it when primer systems evoking the PS-effect are concerned: it increases the length of inverted terminal repeats carried by the non-specifically amplified molecules, which suppresses their amplification even further.

The protocol described here assumes simultaneous RACEs for 3' and 5' flanks, starting from a known fragment of cDNA sequence 500-1000 bases long. This situation is the most common in PCR-based gene hunting. The logistics and schematic protocol of this experiment are outlined on Figure 1. As depicted there, in this case the four gene-specific primers can be designed to allow amplification of the positive control (a part of the known fragment) at the first RACE stage along with nested PCR required at the second RACE stage. If the known fragment is longer than 1000 bases, the RACE primers should be designed to keep 5' and 3' nested steps (see **Fig. 1** for the term definition) within the recommended limits. This may dictate the need to design two additional gene-specific primers (one for each RACE direction) for amplification of positive controls. These additional primers should produce amplicons 300-500 base pairs long when used with their correspondent first-stage RACE primers (*5-1* and *3-1* on Fig. 1).

2. Materials

1. cDNA (amplified or non-amplified, *see* **Note 1)** synthesized according to one of the two methods described in the chapter "*Amplification of representative cDNA samples from microscopic amounts of animal tissue*", pages NN-NN of this book.
2. cDNA dilution buffer: 10 mM Tris-HCl pH 8.0; 10 ng/µL yeast tRNA (*see* **Note 2**).
3. Long-and-Accurate PCR enzyme mix with provided buffer (Advantage2 polymerase mix by Clontech, LA-PCR by Takara, Expand Taq by Boehringer or equivalent).
4. dNTP mix, 10 m*M* each.
5. Oligonucleotides: *see* **Box 1** and **Notes 3 and 4**.
6. Agarose gel (1%) containing ethidium bromide.

3. Methods

3.1. First Stage of RACE

1. Prepare the template by diluting the aliquot of non-amplified cDNA 5-fold in the cDNA dilution buffer. If the use of amplified cDNA is feasible, dilute the aliquot of the product of cDNA amplification 50-fold in the cDNA dilution buffer.
2. Prepare three PCR mixtures, corresponding to 3'-RACE, 5'-RACE and positive control. These mixtures will differ only in primers, so a master mixture without primers can be prepared: 6 μL of 10× PCR Buffer (provided with the polymerase mixture), 1.5 μL of dNTP mix (10 m*M* of each), 3 μL of diluted cDNA template, polymerase mixture sufficient for 60 μL of PCR (see manufacturer's recommendations) and H_2O to 57 μL. Dispense the master mix into three tubes (18 μL to each tube), label the tubes. Add the following primers to the tubes (see Fig. 1 for locations of primer annealing sites):
    a. for **5'-RACE**, 1 μL of 2 μ*M* **5prox** primer and 1 μL of 2 μ*M* fragment-specific primer **5-1**;
    b. for **3'-RACE**, 1 μL of 2 μ*M* **3prox** primer and 1 μL of 2 μ*M* fragment-specific primer **3-1**;
    c. for **positive control**, 1 μL of 2 μ*M* primer **5-1** and 1 μL of 2 μ*M* primer **3-1**.

    Note that the final concentration of primers is 0.05 μ*M*.
3. Perform cycling: for block-controlled thermocyclers, 94 ºC 40 seconds – [annealing temperature of the weakest gene-specific primer, *see* **Note 4**] 1 minute – 72 ºC 2 minutes 30 seconds; for thermocyclers with tube-controlled temperature or simulated tube control, 95 ºC 10 seconds – […] 30 seconds – 72 ºC 2 minutes 30 seconds. Perform as many cycles as required to amplify the known fragment from the particular cDNA. If this value is unknown, do 22 cycles. Check 3 μL of the products on a 1% agarose gel, keeping the PCR tube at room temperature while the electrophoresis runs. If none of the tubes contains detectable product, put the tubes back into thermal cycler and do 5 more cycles. If the products are barely visible, do only 3 more cycles. Repeat the checks on agarose gel and adding cycles according to these guidelines until the products in all three tubes are readily detectable on an agarose gel with standard EtBr staining. However, the total number of cycles should not be more than 40 (*see* **Note 5**).

3.2. Second Stage of RACE

4. Prepare the templates for second stage RACEs by diluting the product of the first stages 50-fold in cDNA dilution buffer.
5. Prepare four PCR mixtures, corresponding to nested/step-out 3'-RACE, nested/step-out 5'-RACE and two negative controls, one for each RACE direction. A master mix for the four samples includes: 8 μL of 10× PCR Buffer (provided with the polymerase mixture), 2 μL of dNTP mix (10 m*M* of each), 4 μL of 2 μ*M* primer *Udist*, polymerase mixture sufficient for 80 μL of PCR (see manufacturer's recommendations) and H_2O to 76 μL. Dispense the master mix into two tubes, 38 μL to each. To one tube add 2 μL of diluted first stage 3'-RACE; to the other - 2 μL of diluted first stage 5'-RACE. Dispense each of the two mixtures into two tubes, 19 μL to each. One of these resulting mixtures will contain the *Udist* primer alone and will serve as a negative control; to the other add 1 μL of 2 μ*M* nested gene-specific primer corresponding to the RACE direction (*3-2* for 3' RACE and *5-2* for 5' RACE). Perform cycling using the same program as described above for the first stage, but with the fewer number of cycles. The number of cycles before the first check on agarose gel should be 12. After that add more cycles if necessary using the guidelines described above; the only difference is that the maximum allowed number of cycles in this case is 20 instead of 40 (*see* **Note 5**).
6. Examine the RACE results on agarose gel. Load the products of first stage side-by side with corresponding second stage products and negative controls from the second stage, achieve a good resolution and analyze the image. Normally the products of RACE, especially if the flanks to be obtained are longer than 1000 base pairs, exhibit multiple bands of different sizes, of which some may be non-specific (*see* **Note 7**). A specific band is the one that does not appear in the negative control (*see* **Note 8**); it also may correspond to a band in the product of the first stage that is longer by the size of the nested step (see **Fig. 1** for the term definition). The latter cannot be expected if the second stage amplification took more than 15 cycles. When the product of RACE turns out to be complex, we recommend picking the specific bands out of the agarose gel, re-amplifying them and cloning one-by-one (*see* **Note 9**). Otherwise, the whole second stage product can be cloned using any vector system suitable for cloning PCR products (such as pGEM-T vector system by Promega), following the manufacturer's protocol.

4. Notes

1. RACE can be performed either starting from amplified cDNA (obtained according to the methods A or B described on pages NN-NN of this book), or from raw non-amplified cDNA. In the case of method A, this raw cDNA is the ligation mixture of double-stranded cDNA with pseudo-double-stranded adaptor, while in the case of method B it is the first-strand synthesis reaction mixture. Using amplified cDNA is highly recommended for short transcripts (total length less than 1500 bases), since there is much less background amplification. For longer transcripts the farthest 5'-flank may be under-represented in the amplified cDNA, so the use of raw non-amplified cDNA is recommended.
2. cDNA dilution buffer contains yeast tRNA to stabilize the DNA in low concentration. The diluted samples can be stored at –20C for several months.
3. Adaptor-specific primers (**Box 1**) are designed to work on cDNAs synthesized with the methods A and B described on pages NN-NN of this book. Note that the primer *5prox* is different for these two methods due to the different 5'-flanking adapter sequence. The adapter-specific primers should be ordered with the highest degree of purification (HPLC or PAAG).
4. Gene-specific primers should be designed following the five simple guidelines.
    a. The 3'-terminal base should be A or T.
    b. The **five** 3'-terminal bases should include no more than two G or C.
    c. The **four** 3'-terminal bases should not find a perfect match within the primer being designed and within primers that are going to be used in PCR with this primer (adapter-specific primers and another gene-specific primer to amplify positive control, if applicable).
    d. The length of the primer should be at least 20 bases.
    e. The annealing temperature of the primer should be equal or higher than 60^{o}C; calculated by the formula: $4x(G+C) + 2x(A+T) + 3$

    In RACE amplification reactions, the annealing should be set to the lowest of the temperatures calculated for the participating gene-specific primers using this formula.
5. If during the first stage of RACE the product in the positive control is not seen after 38-40 cycles, the gene-specific primers were badly designed or synthesized, or the particular cDNA sample does not contain the target transcript. If the positive control shows a well-detectable band, while the RACEs are empty even at 5-8 cycles more than the cycle number required to amplify the positive control, the problem may be the bad quality of adapter-specific primers (*5prox* and *3prox*) or the unsuccessful cDNA synthesis.
6. If no bands discriminating the experimental samples from negative controls are detected after 20 PCR cycles during second stage of RACE, this means either that the first stage amplification did not work

properly or something is wrong with the nested gene-specific primers. Note that although adding more cycles may eventually produce some bands, these most probably would not correspond to the transcript of interest.

7. It is very common for specific RACE products to appear as several bands ("RACE ladders"). In 5'-RACE, this is explained by partial degradation of the original mRNA or fall-offs of the reverse transcriptase during first-strand synthesis. In 3'-RACE, multiple bands correspond to the sites of non-specific annealing of oligo-dT-containing primer within the mRNA rather than at the poly-A tract. In addition to these artifactual sources, there may be natural causes of "RACE ladders" such as presence of multiple splice forms or presence of high-similarity repeats within the target transcript to which the gene-specific primers anneal.
8. The negative control may show some product of amplification, especially when the number of cycles performed during second stage of RACE approaches 20. This product looks like a condensed smear with the fragment lengths from 4 to 8 kb, and represents the longest cDNA fragments that were present in the sample. Fragments that long are affected by PCR-suppression effect the least of all and have a chance to be amplified to the detectable concentration, especially in the absence of amplification of the target product.
9. To pick the band out of agarose gel, stab the gel in the middle of the band with a pipetman tip (cut the tip to make an opening approximately 1 mm wide). Avoid exposing the gel to ultraviolet for prolonged periods of time. Squeeze the piece of agarose out of the tip into 20 μL of milli-Q water and either leave it overnight at 4C or incubate it at 55C for 1 hour; then take one microliter of the liquid to re-amplify the band using the same primers, controls, program and cycle number as at second stage of RACE.

5. References

1. Yu, Y. P., Lin, F., Dhir, R., Krill, D., Becich, M. J. and Luo, J. H. (2001) Linear amplification of gene-specific cDNA ends to isolate full-length of a cDNA. *Analytical Biochemistry* 292, 297-301.
2. Tellier, R., Bukh, J., Emerson, S. U. and Purcell, R. H. (1996) Amplification of the full-length hepatitis A virus genome by long reverse transcription-PCR and transcription of infectious RNA directly from the amplicon. *Proceedings of the National Academy of Sciences of the United States of America* 93, 4370-4373.
3. Shi, X. Z. and Kaminskyj, S. G. W. (2000) 5 ' RACE by tailing a general template-switching oligonucleotide. *Biotechniques* 29, 1192.
4. Matz, M., Shagin, D., Bogdanova, E., Britanova, O., Lukyanov, S., Diatchenko, L. and Chenchik, A. (1999) Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Research* 27, 1558-1560.
5. Kruger, M., Beger, C., Welch, P. J., Barber, J. R. and Wong-Staal, F. (2001) C-SPACE (cleavage-specific amplification of cDNA ends): a novel method of ribozyme-mediated gene identification. *Nucleic Acids Research* 29, U44-U50.
6. Koehler, D. R., Tellier, R., Hu, J. and Post, M. (1999) System for PCR identification of cDNA ends (SPICE). *Biotechniques* 27, 46.
7. Huang, S. H., Hu, Y. Y., Wu, C. H. and Holcenberg, J. (1990) A Simple Method for Direct Cloning Cdna Sequence That Flanks a Region of Known Sequence from Total Rna by Applying the Inverse Polymerase Chain-Reaction. *Nucleic Acids Research* 18, 1922-1922.
8. Frohman, M. A., Dush, M. K. and Martin, G. R. (1988) Rapid Production of Full-Length Cdnas from Rare Transcripts - Amplification Using a Single Gene-Specific Oligonucleotide Primer. *Proceedings of the National Academy of Sciences of the United States of America* 85, 8998-9002.
9. Eyal, Y., Neumann, H., Or, E. and Frydman, A. (1999) Inverse single-strand RACE: An adapter-independent method of 5 ' RACE. *Biotechniques* 27, 656-+.
10. Chenchik, A., Diachenko, L., Moqadam, F., Tarabykin, V., Lukyanov, S. and Siebert, P. D. (1996) Full-length cDNA cloning and determination of mRNA 5' and 3' ends by amplification of adaptor-ligated cDNA. *Biotechniques* 21, 526-534.
11. Lukyanov, K. A., Gurskaya, N. G., Bogdanova, E. A. and Lukyanov, S. A. (1999) Selective suppression of polymerase chain reaction. *Bioorganicheskaya Khimiya* 25, 163-170.
12. Siebert, P. D., Chenchik, A., Kellogg, D. E., Lukyanov, K. A. and Lukyanov, S. A. (1995) An Improved Pcr Method for Walking in Uncloned Genomic DNA. *Nucleic Acids Research* 23, 1087-1088.
13. Shagin, D. A., Lukyanov, K. A., Vagner, L. L. and Matz, M. V. (1999) Regulation of average length of complex PCR product. *Nucleic Acids Res* 27, e23.

Figure legends

Figure 1. RACE logistics and scheme of the protocol. The sequences of the adapter-specific primers are pattern-coded in accordance to Box 1.

Figure 2. Typical example of RACE results. In this experiment, 3' and 5' flanks of cDNA coding for cadherin-related protein expressed in central nervous system of sea hare *Aplysia californica* were obtained. The template was the amplified cDNA prepared from six identified neurons (metacerebral cells), according to method A (see pages NN-NN). Lane **M**: 1 kb DNA ladder (Promega); Lane **1(+)**: positive control from the first step. Three lanes correspond to each RACE direction: **1**: first step, **2**: second step; **2(-)**: negative control from the second step.

Box 1: *Legend to primer sequences*

5prox

(equals DAP-PAP in method A and DAP-TS in method B)

CGACGTGGACTATCCATGAACGCAACTCTCCGACCTCTCACCGA *(method A)*

CGACGTGGACTATCCATGAACGCAAAGCAGTGGTATCAACGCAGAGT *(method B)*

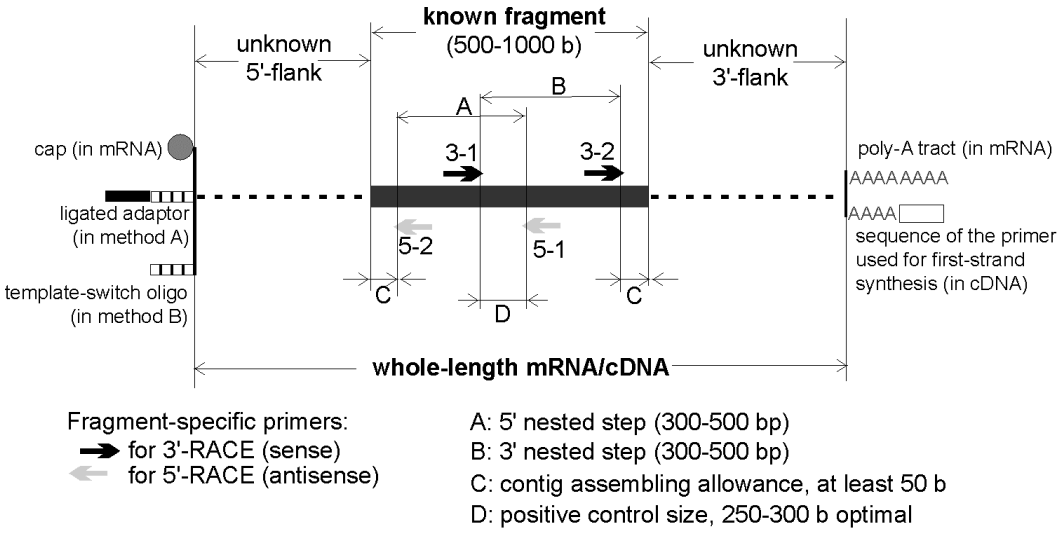
3prox

(equals DAP-Trsa in both cDNA synthesis methods)

CGACGTGGACTATCCATGAACGCACGCAGTCGGTAC (T)13

Udist

TCGAGCGGCCGCCCGGGCAGGT CGACGTGGACTATCCATGAACGCA



	5′-RACE	*3′-RACE*
STEP 1		
template	1 μl of 1/5 dilution of SMART cDNA or adapter-ligated cDNA or (for transcripts up to 2000 b long) 1 μl of 1/50 dilution of amplified cDNA	
primers for target	*5′ proximal primer (5prox)* — *5-1*	*3-1* — TTTTT — *3′-proximal primer (3prox)*
target product	*5prox* — *3-1* — *3-2* — AAAA / TTTT — *3prox* — *5-2* — *5-1*	
control (positive)	*3-1* — *5-1*	
No. of cycles	commonly 20-35 - until the target product is well-visible on the agarose gel (normally 2-4 cycles more than required to amplify positive control)	
STEP 2		
template	1 μl of 1/50 dilution of the **step 1** product	
primers for target	*universal distal primer (Udist)* — *5-2*	*3-2* — *universal distal primer (Udist)*
target product	*Udist* — *5-2*	*3-2* — AAAA / TTTT — *Udist*
control (negative)	primer *Udist* alone	
No. of cycles	commonly 12-15; but **no more than 20**	

Figure 1. RACE logistics and scheme of the protocol. The sequences of the adapter-specific primers are pattern-coded in accordance to Box 1.

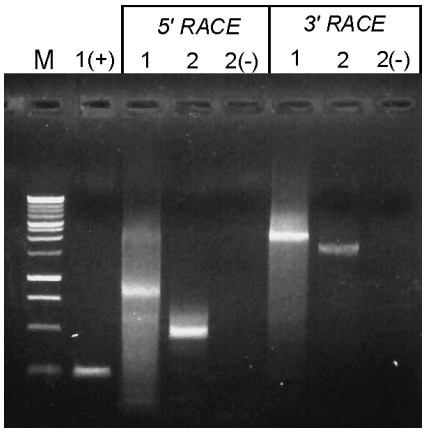


Figure 2. Typical example of RACE results. In this experiment, 3' and 5' flanks of cDNA coding for cadherin-related protein expressed in central nervous system of sea hare *Aplysia californica* were obtained. The template was the amplified cDNA prepared from six identified neurons (metacerebral cells), according to method A (see pages NN-NN). Lane **M**: 1 kb DNA ladder (Promega); Lane **1(+)**: positive control from the first step. Three lanes correspond to each RACE direction: **1**: first step, **2**: second step; **2(-)**: negative control from the second step.