Note

Cloning of Full-Length cDNA of Teleost Corticotropin-Releasing Hormone Precursor by Improved Inverse PCR

Hisato KUNIYOSHI,[†] Yukio FUKUI, and Yoichi SAKAI

Graduate School of Biosphere Science, Hiroshima University, 1-4-4 Kagamiyama, Higashi-hiroshima, Hiroshima 739-8528, Japan

Received February 6, 2006; Accepted April 7, 2006; Online Publication, August 23, 2006 [doi:10.1271/bbb.60067]

We designed a new inverse PCR protocol combined with switching mechanism at 5' end of RNA transcript (SMART) technology, and applied it to the cloning of teleost corticotropin-releasing hormone precursor cDNA. Due to the advantages of both techniques, this method can efficiently amplify the complete 5'- and 3'ends of cDNA in a single reaction, and might prove to be an alternative to the conventional rapid amplification of cDNA ends (RACE) approaches.

Key words: inverse PCR; corticotropin-releasing hormone precursor; teleost; *Halichoeres tenuispinnis*

Rapid amplification of cDNA ends $(RACE)^{1,2}$ is an efficient strategy to obtain sequence information on either 3'- or 5'-regions flanking a known sequence of cDNA, but the conventional RACE method often amplifies non-specific products, which can result from the use of only one gene-specific primer.^{2,3} In such cases, nested PCR is necessary to increase specificity.²

An alternative strategy is inverse PCR.⁴⁾ In this method, double-strand cDNA (ds-cDNA) is circularized by self-ligation and used as a PCR template. With a pair of gene-specific primers arranged in a back-to-back orientation, the cDNA fragment containing both 3'- and 5'-ends can be amplified using a standard PCR protocol (Fig. 1). Inverse PCR has two advantages: (1) non-specific PCR products are less likely to be amplified because of the use of two gene-specific primers,^{3,5)} and (2) both 3'- and 5'-regions are amplified in a single reaction. In spite of these advantages, inverse PCR has not been as popular as RACE.

To improve inverse PCR, we combined it with SMART (switching mechanism at 5' end of RNA transcript) technology,⁶⁾ commercialized by Clontech in their SMART[™] PCR cDNA Synthesis Kit (Fig. 1). SMART technology utilizes the template-switching activity of MMLV-reverse transcriptase to add an



13/

Fig. 1. Schematic Outline of Improved Inverse PCR.

Thick and thin lines represent mRNA and cDNA respectively. Arrows indicate primers. White boxes at the ends of cDNAs represent the anchor sequences introduced with the SMARTTM PCR cDNA Synthesis Kit. The known sequence obtained by degenerate PCR is shaded.

anchor sequence to the 3'-end of newly synthesized first-strand cDNA. In combination with an oligo-dT primer with a specific 5' heel sequence, first-strand cDNA with anchor sequences at both ends is synthesized. The cDNA population is amplified by longdistance PCR (LD-PCR) with a primer, 5' PCR primer

[†] To whom correspondence should be addressed. Fax: +81-82-424-2459; E-mail: hkuni@hiroshima-u.ac.jp

Abbreviations: CRH, corticotropin-releasing hormone; ds-cDNA, double-strand cDNA; LD-PCR, long-distance polymerase chain reaction; RACE, rapid amplification of cDNA ends; SMART, switching mechanism at 5' end of RNA transcript

IIA (kit supplied), which corresponds to the anchor sequences. To make possible the subsequent self-ligation, we used the 5' PCR primer IIA that had been phosphorylated with T4 polynucleotide kinase (Takara, Kyoto, Japan). The resulting ds-cDNA was circularized by self-ligation and then used as the template for inverse PCR.

Utilizing SMART technology, our inverse PCR protocol has additional advantages. Since the templateswitching phenomenon occurs when reverse transcriptase has reached the end of the mRNA template, fulllength first-strand cDNAs but not prematurely terminated ones are tagged with the anchor sequence, and therefore enriched during LD-PCR.⁶⁾ This property is favorable for the cloning of full-length cDNA. Furthermore, only 0.05–1 μ g of the starting total RNA is required. We applied this protocol to the cloning of corticotropin-releasing hormone (CRH) precursor cDNA of a marine wrasse, *Halichoeres tenuispinnis*.

First we obtained a partial sequence of the wrasse CRH precursor cDNA by degenerate PCR. First-strand cDNA was synthesized from wrasse brain total RNA using an RNA PCR Kit (Takara). Based on the amino acid sequences conserved among teleost CRH precursors,⁷⁾ we designed two degenerate primers, sense primer CRH-1 (5'-GGNGARGARTAYTTYATHMG-3') and antisense primer CRH-2 (5'-TGYTCNGCNC-GNGACATYTC-3') (Y; C/T, R; A/G, H; A/C/T, M; A/C, N; A/C/G/T) (Fig. 3). PCR cycling conditions were 30 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72°C for 30s. A 280-bp product encoding the CRH precursor was amplified. On the basis of its nucleotide sequence, a pair of oppositely directed primers (CRH-5A and CRH-3S) was designed for inverse PCR (Fig. 1 and Fig. 3).

Next we performed improved inverse PCR to obtain the sequences of the 5'- and 3'-regions. First-strand cDNA was synthesized from 1 µg of wrasse brain total RNA using the SMART[™] PCR cDNA Synthesis Kit (Clontech, Palo Alto, CA, USA) according to the manufacturer's instructions. Double-strand cDNA was prepared by LD-PCR using PrimeSTARTM HS DNA polymerase (Takara), which can produce blunt-ended ds-cDNA. A PCR reaction was set up in a total volume of 50 µl containing 1 µl of the first-strand cDNA (equivalent to 100 ng of total RNA), 0.8 µM phosphorylated 5' PCR primer IIA, 0.2 mM each dNTP, 1.25 U PrimeSTARTM HS DNA polymerase, and $1 \times$ PrimeSTARTM Buffer (Takara). After an initial 1-min denaturing step at 95 °C, 15 cycles of amplification were performed using a cycle profile of 95 °C for 5 s, 65 °C for 5 s, and 68 °C for 6 min. After reagents and primers were removed with MicroSpin S-400 HR Columns (GE Healthcare, Piscataway, NJ, USA), dscDNA was circularized using a DNA Ligation Kit Ver. 2.1 (Takara). Inverse PCR was carried out in a 50-µl reaction mixture containing 1µl of circularized ds-cDNA, 0.5 µm primers (CRH-5A and CRH-3S),



Fig. 2. Inverse PCR for *H. tenuispinnis* CRH Precursor cDNA. PCR products were electrophoresed on a 2% agarose gel. A 1.0kbp product (indicated by an arrowhead) was amplified from the circularized ds-cDNA (lane C) but not from the linear ds-cDNA (lane L). The sizes of markers are indicated on the left.

0.2 mM each dNTP, 1.25 U Ex Taq Hot Start Version (Takara), and $1 \times Ex Taq$ Buffer (Takara). PCR parameters were 95 °C for 1 min and 35 cycles of 95 °C for 30 s, 67 °C for 30 s, and 72 °C for 2 min, followed by 72 °C for 7 min. As shown in Fig. 2, an approximately 1.0-kbp fragment was amplified from the circularized ds-cDNA but not from the linear ds-cDNA (i.e., before self-ligation). Sequence analysis revealed that the 1.0kbp amplicon contained both 3'- and 5'-regions of the wrasse CRH precursor cDNA. We found heterogeneity in the length of the 3'-untranslated region, which resulted from a difference in the polyadenylation site (Fig. 3, arrowheads). A 250-bp fragment is likely to have been a non-specific product because it was amplified from linear ds-cDNA as well as circularized ds-cDNA (Fig. 2).

Finally, full-length cDNA was amplified by end-toend PCR with primers, CRH-UP and CRH-DW, corresponding to the 5'- and 3'-ends of cDNA respectively (Fig. 1 and Fig. 3). First-strand cDNA prepared using the RNA PCR Kit (Takara) was used as a template. To suppress errors in the polymerase reaction, end-to-end PCR was carried out using the Expand High Fidelity^{PLUS} PCR System (Roche, Basel, Switzerland). PCR parameters were 35 cycles of 94 °C for 10 s, 55 °C for 30 s, and 72 °C for 1 min. A 1.0-kbp product was amplified. Since Northern blot analysis probed with wrasse CRH precursor cDNA detected an approximately 1.0-kb tran-

Improved Inverse PCR for Cloning of Full-Length cDNA

GA	CRH-UP GAAAGTAACTATATACTTGTTTCCTCAAAGAAGTGAAGGAGGGCGGCTTCTCGCCAACTG															60				
АI	ATACCTGGCAACCTACGCGGCTGTGCTGGACCTCCTGTAGAGACTGAGATTCCTAGATAT															120				
CC	TGA	CAT	gaa	GCT	CAA	TTT.	ACT	TGG	AAC	CAC	CGT	GAT	TCT	GCT	AGT	TGC	CTT	CTT	ACCC	180
		М	ĸ	Ц	N	Ц	Ц	G	т	т	V	1	Ц	Ц	V	A	F.	Ц	Р	18
CG	CTA	CGA	ATG	TCG	GGC	TAT	TGA	GAG	CCC	TGG	CGG	TGC	CCT	GCG	CGT	CCC	AGC	TCC	CCAA	240
R	Y	Ε	C	R	A	I	Е	S	Ρ	G	G	A	L	R	v	Ρ	A	Ρ	Q	38
GC	CCA	AAA	CTC	CCA	GCA	GCA	GCA	.GCA	GCA	GCA	GTC	TGG	TCC	CAT	ССТ	GGA	GCG	TCT	GGGT	300
A	Q	N	S	Q	Q	Q	Q	Q	Q	Q	S	G	P	I	L	Е	R	L	G	58
CRH-5A GAGGAGTATTTCATCCGACTGGGCAACGGGGACTCTAACTCTTTCCCATCTTCGTCCATG															360					
E	E	Y	F	I	R	L	G	N	G	D	S	N	S	F	P	s	s	S	M	78
TA	TCC	TCC	TTC	CCT	CTA	CAA	CAA	AGC	GCT	GCA	ACT	CCA	.GCA	.GCT	GAC	GCG.	ACG	TCT	TTTA	420
Y	Р	Р	S	Г	Y	N	ĸ	A	Г	Q	Г	Q	Q	Г	т	R	R	Г	Г	98
CA	AGG	GAA	AGT	TGG	AAA	CAT	CAG	AGC	GCT	CAT.	AAG	CGG	CTT	CGG	AGA	CCA	AGG	AGA	CGAG	480
Q	G	к	v	G	N	I	R	A	L	I	s	G	F	G	D	Q	G	D	Е	118
_				CRF	I-35	3														
ΠC		aaz.	aza	aaa	770	and.	ama		~					aam		_ ~ ~	~ ~	~	aara	
10	GAI	GGA	GAG	GGG	AAG	GAG	GTC	CGA	GGA	CCC	GCC	GAT	'A'I'C	CCT	GGA	TCT	GAC	C'I''I'	CCAC	540
s	M	E	GAG R	GGG	R R	R	s s	E	GGA D CRH	<u>CCC</u> P - 2	GCC P	GAT I	s S	L	D GGA	L L	GAC T	C'1''1' F	H	540 138
S CI	M	E CCG	R R GGA	GGG GAT	R GAT	R GGA	GTC S GAT	E GTC	GGA D CRH CAA	<u>CCC</u> P -2 GGC	GCC P GGA	GAT I ACA	S GCT	L GGC	D TCA	L GCA	GAC T GGC	GCA	H AAAT	540 138 600
S CI	M M GCT L	E CCG R	R R GGA E	GGG G GAT M	AAG R GAT M	R R GGA E	GTC S GAT M	E GTC	GGA D CRH CAA K	CCC P -2 GGC A	GCC P GGA E	GAT I ACA Q	GCT	L GGC A	D D TCA	GCA	GAC T GGC A	GCA Q	H AAAT N	540 138 600 158
S CI L	M GCT L	E CCG R	R R GGA E	GGG G GAT M	R R GAT M	R GGA E	GTC S GAT M	GTC	GGA D CRH CAA K	<u>CCC</u> P -2 GGC A	GCC P GGA E	GAT I ACA Q	GCT L	L GGC A	GGA D TCA Q	GCA	GAC T GGC A	GCA Q	H AAAT N	540 138 600 158
CI L	M GCT L	E CCG R AAG	R GGA E AAT	GGG G GAT M GAT	R GAT M GGA	R GGA E GCT	GAT M CTT	CGA E GTC S	GGA D CRH CAA K AAA	CCC P -2 GGC A ATG	GCC P GGA E AAT	GAT I ACA Q TTC	GCT L	EGGC A TCC	GGA D TCA Q GCC.	GCA Q AAA	GAC T GGC A GAT	GCA Q TTC	H AAAT N ACTT	540 138 600 158 660
CT L AA	M CGCT L CAG R	E CCG R AAG R	R GGA E AAT M	GGG GAT M GAT M	R GAT M GGA	R GGA E GCT L	GTC S GAT M CTT F	CGA E GTC S CGG G	GGA D CRH CAA K AAA K	CCC P -2 GGC A ATG	GCC P GGA E AAT	GAT I ACA Q TTC	GCT L	EGGC A	<u>GGA</u> D TCA Q GCC	GCA Q AAA	GAC T GGC A GAT	GCA GCA Q TTC	H AAAT N ACTT	540 138 600 158 660 168
CTI L AA N GA	M <u>GCT</u> L CAG R ATTT	E CCG R AAG R TCT	R GGA E AAT M GTT	GGG GAT M GAT M CTT	R GAT M GGA GGA E CTC	R GGA E GGA E CTT	GTC S GAT M CTT F TTC	CGA E GTC S CGG G CGG G	GGA D CRH CAA K AAA K CTT	CCC P -2 GGC A ATG ATG	GCC P GGA E AAT TCA	GAT I ACA Q TTC TTT	GCT CAC	ECCT L CGGC A TCC	GGA D TCA GCC. TCA	GCA GCA Q AAA	GAC T GGC A GAT CAA	GCA GCA Q TTC	H AAAT N ACTT TGCT	540 138 600 158 660 168 720
CT CT L AA N GA CT	M <u>GCT</u> L CAG R TTT GTA	E CCG R AAG R TCT CAG	R GGA E AAT M GTT TAT	GGG GAT M GAT M CTT AGT	R GAT M GGA E CTC GCT	R GGA E GGA E GCT CTT GCT	GTC S GAT M CTT F TTC TTC	CGA E GTC S CGGG G TTTT TTTT	GGA D CRH CAA K AAA K CTT CTC	CCC P -2 GGC A ATG ATG TAT	GCC P GGA E AAT TCA TCA	GAT I ACA Q TTC TTT TTA	GCT CAC	CCT GGC A CCA CCA	D TCA Q GCC. TCA TAA	CCG	GAC T GGC A GAT CAA	GCA GCA Q TTC ACA	AAAT N ACTT TGCT ATGG	540 138 600 158 660 168 720 780
CT CT AAA N GAA CT AG	M GCT L CAG R TTT GTA	E CCG R AAG R TCT CAG	R <u>GGA</u> <u>E</u> <u>AAT</u> <u>M</u> GTT TAT CGGG	GGG G GAT M GAT M CTT AGT CTC	R GAT M GGA CTC GCT GGC	R GGA E GCT L CTT GCT	GAT GAT M CTT F TTC TTA	CGGA E GTC S CGGG G TTTT TCA	GGA D CRH CAA K AAAA K CTT CTC TCC	CCC P -2 GGC A ATG. * CTT TAT GAT	GCC P GGA E AAAT TCA TCA TAT	GAT I ACA Q TTC TTT TTA	GCT GCT CAC CAC TAG	GGC A TCC CCA CCA CTT	D TCA Q GCC. TCA TAA TTT	I GCA Q AAAA CCG CCT TGA	GAC T GGC A GAT GAT CAA CAA	CTT F GCA Q TTTC ACA ACA	H AAAT N ACTT TGCT ATGG TGTC	540 138 600 158 660 168 720 780 840
CT L AAA GAA CT ACC AAA	M <u>GCT</u> <u>L</u> <u>CAG</u> R TTTT GTA TGT ATC	E CCCG R AAG R TCT CAG AAA TGT	R GGAA E AAAT M GTT TAT CGGG GAA	GGG G <u>GAT</u> M GAT M CTT AGT CTC TTG	R GAT M GGA E CTC GGC GGC CAC	R GGA E GCT L GCT TTG CGT	GAT GAT M CTTT F TTC TTAA TCT	CGA E GTC S CGGG G TTT TCA TCA	GGA D CRH CAA K AAAA K CTT CTC TCC	CCC P -2 GGC A ATG. * CTT TAT GAT	GCC P GGA E AAAT TCA TAT TGT. AGA	GAT I ACA Q TTC TTT TTA ACC TAA	GCT GCT CACCAC CACCAC TAG TTG AACC	L GGCCA TCCCA CCCA GCTT GCCA	GGA D TCA GCC. TCA TTA TTT TTT	TCT L GCA Q AAAA CCCG CCT TGA GGA	GAC T GGC A GAT CAA CAA CAA TGT TTA	CTT F GCA Q TTTC ACA ACA TGG	AAAT N ACTT TGCT ATGG TGTC ATAG	540 138 600 158 660 168 720 780 840 900
CT AAA CT GA CT AC AC	M GCT L CAG R TTT GTA TGT ATC CAT	E CCG R AAG R TCT CAG AAA TGT TGC	R GGA E AAT M GTT TAT CGG GAA AAT	GGG G GAT M GAT CTT CTT CTT CTC CTC GAC	AAG R GAT M GGA CTC GGC GGC CAC CAC	R GGA E GCT L CTT GCT TTG CGT	GTC S GAT M CTT F TTC TTC TTA TCT TGC	CGGA E GTC S CGGG G TTTT TCA TGA TCA	GGA D CRH CAA K AAA K CTT CTC TCC AGT	CCCC P -2 GGC A ATG. * CTT TAT GAT TTG. TTT	GCC P GGA E AAAT TCA TGT. TGT. AGA	GAT I ACA Q TTTC TTTT TTTA ACCC TAA	ATC S GCT CAC CAC CAT	CCCT CCCA	GGA D TCA Q GCC. TCA TTA TTT TTT TTT	TCT L GCA Q AAAA CCCG CCT TGA GCA TGA	GAC T GGC A GAT CAA CAA CAA TGT TTA GAT.	CTT F GCA Q TTTC ACA ACA TGG AAAC	AAAT N ACTT TGCT ATGG TGTC ATAG	540 138 600 158 660 168 720 780 840 900 960
CT <u>AAA</u> CT AC AAA AC	M GGCT L CAG R TTTT GTA TGT ATC CAT	E CCG R AAG R TCT CAG AAA TGT TGC	GAG R GGA E AAAT CGGG GAA AAAT	GGG G G GAT M GAT CTT CTT CTC CTC CTC GAC	R GAT M GGA CTC GGC CTC GGC CAC CAC	R GGA E GCT L CTT GCT TTG CGT ACT	GTC S GAT M CTT F TTC TTC TTA TCT	CGGA E CGGG CGGG CTTT TTT TCA TCA	GGA D CRH CAA K AAAA K CTT CTC TCC AGT	CCC P -2 GGC A ATG * CTT TAT GAT TTG TTT	GCC P GGA E AAAT TCA TCA TGT. AGA TAGA	GAT I ACA Q TTC TTT TTA ACC TAA	ATC S GCT CAC CAC TAG TAG AAC	GCCT CCCA	GGA D TCA Q GCC. TCA TTA TTT TTT ATT	TCT L GCA Q AAAA CCG CCT TGA GGA TGA	GAC T GGC A GAT CAA CAA CAA TGT TTA GAT	CTT F GCA Q TTTC ACA ACA TGG AAC	AAAT N ACTT TGCT ATGG TGTC ATAG TATG	540 138 600 158 660 168 720 780 840 900 960
CT L AP R CT AC AC AC AC TT	M <u>GCT</u> <u>L</u> <u>CAG</u> R TTTT GTA TGT CAT TGT	E <u>CCG</u> <u>R</u> <u>AAG</u> TCT CAG AAA TGT TGC T <u>AA</u>	R GGA E AAAT M GTT TAAT CGG GAA AAAT TAA	GGG G M GAT CTT AGT CTC GAC GAC	R G <u>GAT</u> M G <u>GGA</u> CTC GGCT GGCT CAC CAC CAC	R GGA E GCT L CTT GCT TTG CGT ACT DW GTG	GTC GAT M CTT F TTC TTA TCT TGC CAA	CGGA E GTC S CGGG G TTTT TCA TCA TCA TCA	GGA D CRH CAA K AAAA K CTT CTC CTC CTC TCC AGT	CCC P -2 GGC A TAT CTT TAT GAT TTT ATT	GGA P GGA E AAT TCA TAT TGT. AGA TAG. TTC	GAT I ACA Q TTC TTT TTT ACC TAA AAT TGT	ATC S GCT CAC GTA TAG TTG CAT	CCCT L GGC TCC CCA CCA GCA CCTT GCA	GGA D TCA Q GCC. TCA TTA TTT TTT ATT	ICTA I GCA Q AAAA CCG CCT TGA GGA TAT GAA	GAC T GGC A GAT CAA CAA TGT TTA GAT	CTT F GCA Q TTC ACA ACA TGG AAAC AAAC	AAAAT N ACTT TGCT TGCT AATGG TATGG TATGG	540 138 600 158 660 168 720 780 840 900 960

Fig. 3. Nucleotide and Deduced Amino Acid Sequences of the Longest *H. tenuispinnis* CRH Precursor cDNA. The putative signal peptide is underlined. The amino acid sequence for CRH mature peptide is boxed. Broken and solid arrows indicate the degenerate and gene-specific primers respectively. Putative polyadenylation signal sequences in the 3'-untranslated region are boxed. Arrowheads indicate the polyadenylation sites of shorter transcripts.

script (data not shown), it is probable that the amplicon included the full-length sequence of CRH precursor mRNA.

As shown in Fig. 3, the wrasse CRH precursor cDNA (accession no. DQ073097) encoded a 168-amino-acid protein. Similarly to CRH precursors from other species, the wrasse CRH precursor is composed of a putative signal peptide, a cryptic region, and the CRH mature peptide. The CRH mature peptide is located between a potential proteolytic cleavage signal (Arg-Arg) and a C-terminal amidation signal (Gly-Lys).

In this report, we indicate that inverse PCR combined with the SMART system is a useful method, comparable to RACE. So far, we have succeeded in cloning the cDNAs of several neuropeptides from various teleost species (unpublished results). We believe that our inverse PCR protocol should prove to be a powerful tool for cloning of full-length cDNAs.

Acknowledgments

We thank the members of Biology of Aquatic Resources of Hiroshima University for their assistance during fieldwork, and S. Ohtsuka, S. Iwasaki, Y. Takagi, and T. Fumoto of the Takehara Marine Science Station of Hiroshima University for their help in conducting the experiments. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

- Frohman, M. A., Dush, M. K., and Martin, G. R., Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA*, **85**, 8998–9002 (1988).
- Schaefer, B. C., Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. *Anal. Biochem.*, 227, 255–273 (1995).
- Friedrich, M., Grahnert, A., and Hauschildt, S., Analysis of the 3' UTR of the ART3 and ART4 gene by 3' inverse RACE-PCR. *DNA Seq.*, 16, 53–57 (2005).
- 4) Huang, S. H., Hu, Y. Y., Wu, C. H., and Holcenberg, J., 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. *Nucl. Acids Res.*, **18**, 1922 (1990).

- Maruyama, I. N., Rakow, T. L., and Maruyama, H. I., cRACE: a simple method for identification of the 5' end of mRNAs. *Nucl. Acids Res.*, 23, 3796–3797 (1995).
- Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R., and Siebert, P. D., Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques*, **30**, 892–897 (2001).
- 7) van Enckevort, F. H., Pepels, P. P., Leunissen, J. A., Martens, G. J., Wendelaar Bonga, S. E., and Balm, P. H., *Oreochromis mossambicus* (tilapia) corticotropin-releasing hormone: cDNA sequence and bioactivity. *J. Neuroendocrinol.*, **12**, 177–186 (2000).

1986