Published in final edited form as:

Anal Biochem. 2012 May 15; 424(2): 140–141. doi:10.1016/j.ab.2012.02.030.

Molecular cloning of large alternative transcripts based on comparative phylogenetic analysis and exploration of an EST database

Jiabing Ji^{1,2,*} and Rongfu Wang¹

¹Center for Cell and Gene Therapy, Baylor College of Medicine, Houston, Texas 77030, USA

Abstract

In animals, a gene may be 50 Kb or over and contain multiple alternative transcripts with sequences that are not experimentally validated. Under these special circumstances, PCR based cloning may become very difficult. Here a simple cloning strategy is described using the mNLRC5 gene as an example. We performed comparative phylogenetic analysis between murine and human NLR protein families to anchor the translation start codon, searched an EST database with the 3' end of the genomic DNA sequence to obtain ESTs from the farthest 3' end of the gene, and isolated the full length CDS of the mNLRC5 of about 6 Kb through conventional RT-PCR and 3'RACE.

Keywords

Alternative transcript; Cloning; PCR; RACE; EST; BLAST; BLAT; Phylogenetic analysis

Alternative 3' processing and polyadenylation can produce multiple transcripts from the same gene which is very common in eukaryotes [1]. To date, gene expression and functional analysis still require the isolation and cloning of alternative transcripts. When the gene is very large in size, and the sequences of multiple alternative transcripts are not experimentally validated, PCR based cloning may become very difficult due to its possible bias toward short products. However, coding sequences (CDS) of many homologues from mouse and human are very conserved and comparative phylogenetic analysis of the closest paralogues should provide some important clues. On the other hand, EST sequencing projects (and recently RNA-Seq) [2–3] have generated huge amounts of sequence information, which has not been fully exploited in functional genomics research. We could utilize this rich resource for capturing different 3' ends from a large gene and facilitating the final cloning of full length cDNA or CDS. This paper reports how we combined these methods with conventional RT-PCR and 3' RACE to isolate the full length CDS of about 6 Kb from an alternative transcript of mNLRC5.

The initial attempts for cloning mNLRC5 were based on the predicted sequence deposited in GenBank (mRNA accession: XM_911402) (Protein accession: XP_916495). All RT-PCR starting from the translation start codon at the 5' end failed and the 3'RACE product was

*Corresponding author: ji6@purdue.edu.

²Current address: Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana 47907, USA

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

always shorter. However, when the primers designed from the internal parts were used to investigate the expression pattern in various tissues of mouse, as shown in Supplementary Fig. 1, this gene is expressed in all the tissues examined. This result suggests that the failure to clone the 5' end of the gene could be due to unsuccessful full length first cDNA synthesis and/or wrong sequence prediction. Since human NLRC5 (hNLRC5) has been predicted and experimentally cloned, sequence comparison between hNLRC5 and XP_916495 may differentiate the two possibilities. As shown in Supplementary Fig.2, XP_916495 contains extra amino acids in the N terminus but lacks a segment in the C terminus. If the sequence difference between hNLRC5 and the putative mNLRC5 is true, then we should expect that it may also hold true for other members in this NOD-like protein family which is involved in innate immune response [4-6]. However, multiple alignments [7] of other members including CIITA, NOD1, NOD2, NLRC3 and NLRX1 [8] in the clade containing NLRC5 [9] does not support this notion. Instead, it suggests that the sequence length of human and mouse heterologues in this clade at the protein level are the same between mouse and human in the very N and C termini, except for NOD3. Based on this analysis, it is reasoned that the originally predicted sequence may not be correct or it is a very rare transcript. New primers (m66F3B and m66R5) as shown in Table 1 were designed from the translation start codon (i.e. it begins from the same position as in hNLRC5 and corresponds to the nucleotides in XM_911402) predicted from this alignment and successfully used to isolate the 5' end of mNLRC5 CDS. The PCR product was about 2.2 Kb, cloned into pCDNA3.1Z with a HA tag after double digestion with *EcoRV* and *XbaI*, and sequenced.

The shorter 3' end from the 3' RACE was most likely caused by RT-PCR bias toward short alternative transcripts, although it is also possible that the longer transcripts are expressed at a lower level. We reason that this gene is expressed abundantly in multiple tissues and if this gene contains longer alternative transcripts, it is likely that the farthest 3' end of the longer transcripts could be recovered in the huge EST database. To test this idea, about 15 Kb genomic DNA sequence including the 3' end of mNLRC5 and its extended 3' flanking fragment after successive BLAT [10] against Mouse Assembly (July 2007, NCBI37/mm9) with XM_911402 and the recovered flanking sequences was downloaded from UCSC Genome Bioinformatics (<http://genome.ucsc.edu/>), and used to BLAST against the EST database at NCBI. As expected, several clones were recovered from this attempt, and one of them was BG916856, which was derived from the infiltrating ductal carcinoma of a 5 months old female mouse (Strausberg 1999, NCBI). For the next step, two nested reverse primers (m67R3 and m67R4) based on this EST were designed to verify together with the primer (m67F4) from the middle of mNLRC5 that this EST is really expressed and part of the transcript with cDNA from a mixture of tissues thought to increase the chance of successful amplification. After sequencing, this EST was confirmed as part of the mNLRC5 transcript, and two rounds of new 3' RACE was performed successfully as shown in Fig. 1. Gel purification and sequencing of the second 3' RACE product indicates that it contains a 3' UTR of 1197 bp long. New primers (m67F2 and m67R8) were designed to amplify the 3' part of mNLRC5 CDS, approximately 3.7 Kb of which was successfully ligated after double digestion with *HapI* and *XbaI* into the vector pCDNA3.1Z containing a HA tag already fused in frame with the 5' end of mNLRC5 CDS, and sequenced.

The full length CDS of mNLRC5 is 5748 bp (GenBank accession FJ889356) which encodes 1915 amino acids. After the transfection of HEK293 cells with this plasmid, it was expressed at the predicted size as shown in Supplementary Fig. 3, and subsequently employed successfully in other functional analyses [11].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The author thanks The Wang Lab for financial support and Dr. Jun Cui for kind help with Western Blotting.

References

1. Rampias TN, Fragoulis EG, Sideris DC. Efficient cloning of alternatively polyadenylated transcripts via Hybridization Capture PCR. *Curr Issues Mol Biol.* 2012; 14:1–8. [PubMed: 21555788]
2. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science.* 1991; 252:1651–1656. [PubMed: 2047873]
3. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The Transcriptional Landscape of the Yeast Genome Defined by RNA. *Sequencing Science.* 2008; 320:1344–1349.
4. Chen G, Shaw MH, Kim YG, Nunez G. NOD-like receptors: role in innate immunity and inflammatory disease. *Annu Rev Pathol.* 2009; 4:365–398. [PubMed: 18928408]
5. Inohara N, Chamaillard M, McDonald C, Nunez G. 2005 NOD-LRR proteins: role in host-microbial interactions and inflammatory disease. *Annu Rev Biochem.* 2005; 74:355–383. [PubMed: 15952891]
6. Ting JP, Kastner DL, Hoffman HM. CATERPILLERS, pyrin and hereditary immunological disorders. *Nat Rev Immunol.* 2006; 6:183–195. [PubMed: 16498449]
7. Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 1988; 16:10881–90. [PubMed: 2849754]
8. Ting JP, Lovering RC, Alnemri ES, Bertin J, Boss JM, Davis BK, Flavell RA, Girardin SE, Godzik A, Harton JA, Hoffman HM, Hugot JP, Inohara N, Mackenzie A, Maltais LJ, Nunez G, Ogura Y, Otten LA, Philpott D, Reed JC, Reith W, Schreiber S, Steimle V, Ward PA. The NLR gene family: a standard nomenclature. *Immunity.* 2008; 28:285–287. [PubMed: 18341998]
9. Tattoli I, Carneiro LA, Jehanno M, Magalhaes JG, Shu Y, Philpott DJ, Arnoult D, Girardin SE. NLRX1 is a mitochondrial NOD-like receptor that amplifies NF-kappaB and JNK pathways by inducing reactive oxygen species production. *EMBO Rep.* 2008; 9:293–300. [PubMed: 18219313]
10. Kent WJ. BLAT - the BLAST-like alignment tool. *Genome Res.* 2002; 12:656–64. [PubMed: 11932250]
11. Cui J, Zhu L, Xia X, Wang HY, Legras X, Hong J, Ji J, Shen P, Zheng S, Chen ZJ, Wang RF. NLRC5 negatively regulates the NF-kappaB and type I interferon signaling pathways. *Cell.* 2010; 141:483–496. [PubMed: 20434986]

Figure 1. RT-PCR and 3'RACE for cloning mNLRC5. Primers m67F10 and 3RACEP1, and m67F11 and 3RACEP2 were used for the 1st and 2nd rounds of 3' RACE. The 2nd 3' RACE product (marked as an arrow) was gel purified and sequenced directly. Lane 1 in A and B is 1 Kb ladder from Invitrogen.

Table 1

The oligonucleotides used in this study

name	primer sequence	primer use
m66F3B	TAGATATCATGGACGCTGAGAGCATCCGACTGAA	5' end of mNLRC5 CDS
m66R5	GCTCTAGATGATGCCTTGGGCTGTAATC	5' end of mNLRC5 CDS
m67F4	AGCAGGGACGCACTAGAACCAA	EST hit verification
m67R3	GAGTCATTTCGGCTTAGTCCTTGA	EST hit verification
m67R4	TCCTGTTCCTGAGAGGTCAA	EST hit verification
m67F10	TCAGCCAAGTGAGCACAAAGA	1 st 3' RACE of mNLRC5
RACEP1	GCGAGCACAGAATTAATACGACT	1 st 3' RACE of mNLRC5
m67F11	TTCAAGGACTAAGCCGAATG	2 nd 3' RACE of mNLRC5
RACEP2	GAATTAATACGACTCACTATAGG	2 nd 3' RACE of mNLRC5
m67F2	AAAAGCAGAAAGTGTGGGGATG	3' end of mNLRC5 CDS
m67R8	GCTCTAGATTAAAGAGTCTGCTGGTCAGTGATGGAGA	3' end of mNLRC5 CDS

Note that the underlined bases represent the restriction sites added to facilitate the directional cloning of PCR products.