Acta Biochim Biophys Sin 2014, 46: 997–1000 | © The Author 2014. Published by ABBS Editorial Office in association with Oxford University Press on behalf of the Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. DOI: 10.1093/abbs/gmu084.
Advance Access Publication 18 September 2014

Lab Note

Cloning and bioinformatics analysis of a full-length cDNA of porcine CR1-like gene

Jing Cheng[1], Junbing Jiang[1], Junxing Zhao[1], Zhirui Wang[2,3], Yaogui Sun[1], Haili Ma[1], Kuohai Fan[1], Wei Yin[1], Na Sun[1], Zhiwei Wang[1], Xin Zhao[1], and Hongquan Li[1]*

[1]College of Animal Science and Veterinary Medicine, Shanxi Agricultural University, Taigu 030801, China
[2]Transplantation Biology Research Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02148, USA
[3]MGH-DF/HCC Recombinant Protein Expression and Purification Core, Boston, MA 02148, USA
*Correspondence address. Tel/Fax: +86-354-6288409; E-mail: livets@163.com

Human erythrocyte complement receptor type 1 (CR1), also known as CD35, is a large single-chain transmembrane glycoprotein and an immune adherence receptor of complement component 3b/4b [1]. Its amino acid composition and genetic structure have been reported [2]. In primate, when the erythrocytes bounded with immune complexes (ICs) moved through the liver and spleen, only the ICs were removed by sinusoidal macrophages and the erythrocytes were allowed to return to the circulation [3]. CR1 has received many studies in many species in recent years due to its important role in mammalian immune system activation [4,5]. However, up to date, porcine CR1-like protein is rarely reported. Our previous study proved the existence of immune adherence (IA) between the porcine erythrocytes and complement-opsonized *Escherichia coli* carrying green fluorescent protein gene (GFP-*E. coli*), but the mechanisms still remain unclear, and we obtained the partial CR1-like gene by reverse transcription-polymerase chain reaction (RT-PCR) and rapid amplification of cDNA 3′-end (3′-RACE) (157 bp and 578 bp) (GenBank accession No. JX033989) [6]. We also identified that the erythrocytes of chickens and mice possessed IA function [7]. Based on those findings, the current experiment was designed to amplify the porcine CR1-like gene, carry out bioinformatics analysis with an attempt to obtain its full-length cDNA, and predict the biological properties of porcine CR1-like protein.

All pigs used in this study were Juvenile healthy Landrace (20 ± 2 kg), obtained from commercial pig farm. The animals were raised in compliance with the International Guiding Principles for Biomedical Research Involving Animals (http://www.cioms.ch/frame.1985.texts.of.guidelines.htm), housed under room temperature with good ventilation hygienic conditions and fed *ad libitum*. Pigs were euthanized and bone marrows were sampled and snapped frozen in liquid nitrogen for further use. Total RNA was isolated from bone marrow using Trizol Reagent (Invitrogen, Carlsbad, USA) according to the manufacturer's protocol and was subject to reverse transcription using primeScript®II RT Enzyme Mix (TaKaRa, Dalian, China). The primers used in the present study were synthesized by BGI·Tech (Shenzhen, China) (**Supplementary Table S1**). PCR amplifications were then performed. The 3′- and 5′-end sequences of CR1-like cDNA were obtained by rapid amplification of cDNA end (RACE) using 3′-Full RACE Kit (TaKaRa) and SMARTer® RACE cDNA Amplification Kit (Clontech, Palo Alto, USA). Based on the sequence assembled, the gene specific primers 3′-C1, 3′-C2, and 5′-N1 were designed to pull out the full-length of porcine CR1-like cDNA. For 3′-RACE, the first-strand cDNA was synthesized by using the adaptor in the kit. The 3′-end of porcine CR1-like gene was amplified by nested PCR using the primers 3′-C1 and 3′-C2 with adaptor primer (AP) 1 and AP2. For 5′-RACE, the first-strand cDNA was synthesized by using a modified lock-docking oligo(dT) primer and the SMARTer II A oligonucleotide and SMART
Scribe Reverse Transcriptase in the kit. The 5′-end of porcine CR1-like gene was carried out by touchdown PCR using primers 5′-N1 and UPM. All PCR products were gel-purified using E.Z.N.A.™ Gel Extraction Kit (OMEGA, Norcross, USA) and sub-cloned for sequencing, and all the sequenced PCR fragments were assembled. The full-length cDNA sequence was obtained and submitted to GenBank (accession No. KF286608). Open reading frame (ORF) and amino acid sequence of porcine CR1-like cDNA were deduced by using the Open Reading Frame Finder (http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi) and the DNAMAN software. As shown in **Supplementary Fig. S1**, the full-length cDNA was 4391 bp long, including a 5′-untranslated region (UTR) of 37 bp, a single ORF of 3996 bp, which encodes 1331 amino acids with a molecular weight of 146.6 kDa, and a 3′-UTR of 358 bp, containing a signal sequence of AATAAA for adding the polyA tail at 318 bp downstream of the termination codon. Pig BLAST-like Alignment Tool (BLAT) (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) was used to analyze the exons and the introns. Comparison of the cDNA sequence

with the genomic sequence revealed that the porcine CR1-like genomic DNA located in chromosome 9 and consisted of 26 exons and 25 introns. We first obtained the sequence from one pig, and the other four independent experiments were used to identify the sequence. There is no difference among these sequences.

CR1-like protein domains were predicted by Simple Modular Architecture Research Tool (SMART) (http://smart.embl-heidelberg.de/). Porcine CR1-like protein contained 19 complement control repeat (CCP) domains, which was the hallmark of the regulators of complement activation (RCA) family. All of the 19 domains were indicated by the box in red (**Supplementary Fig. S1**). CCP, also named short consensus repeat or sushi domain, has a length of ~60 amino acids, and it is a conserved protein–protein interaction domain in many proteins [8]. These domains contain a consensus pattern of four invariant cysteine residues and some additional conserved residues. Proteins with CCP domain usually participate in the regulation of complement system and blood coagulation as well [9]. The extracellular portions of porcine CR1-like protein contained 19 CCPs, providing multiple potential sites for C3b recognition. In the porcine CR1-like protein, Cys(I)–Cys(II)–Cys(III)–Trp–Cys(IV) motifs were found in every CCP, which were necessary for the formation of CCP and responsible for its function as an IA receptor protein. Those conserved motifs might contribute to a hydrophobic core within each module, and each domain contained a hydrophobic core wrapped in β-sheets [10]. In human, CCP1–3 are necessary for the binding of C4b, and CCP15–17 can bind to C3b/C4b [11]. We found that the CCP1–3 and 15–17 were conserved between human and pig, suggesting that porcine CR1-like may had same function as human CR1.

We also compared the amino acid sequence of porcine CR1-like protein with other mammalian species by using DNAMAN software (**Supplementary Fig. S2**). The multiple sequence alignment analysis showed that the deduced amino acid sequence of porcine CR1-like protein shared 55% similarity with that of Gorilla (GenBank accession No. XP_004028363.1), 54% similarity with *Homo sapiens* (GenBank accession No. NP_00564.2), 56% similarity with *Macaca mulatta* (GenBank accession No. XP_002801991.1), 62% similarity with *Nomascus leucogenys* (GenBank accession No. XP_003273038.2), 55% similarity with *Pan troglodytes* (GenBank accession No. NP_001180604.1), 67% similarity with *Tursiops truncatus* (GenBank accession No. XP_004329262.1), 38% similarity with *Mus musculus* (GenBank accession No. NP_038527.2), 42% similarity with *Rattus norvegicus* (GenBank accession No. NP_001005265.1), 58% similarity with horse (GenBank accession No. XP_005609892.1), 54% similarity with *Oryctolagus cuniculus* (GenBank accession No. XP_002717475.1), 57% similarity with dog (GenBank accession No. XP_005622374.1), and 36% similarity with *Bos taurus* (GenBank accession No. ADO51744.1). To further determine the evolutionary relationship between porcine CR1-like protein and other species, the CR1-like protein sequences were used to construct the phylogenetic tree based on the multiple amino acid sequences generated by ClustalX. The phylogenetic tree was aligned using the neighbor-joining method and constructed by MEGA 5.2. As shown in **Supplementary Fig. S3**, porcine CR1-like protein locates at a relatively individual position, which belongs to the mammalian artiodactyla. Interestingly, pig shares close homology with horse and *T. truncates*, but not *B. taurus* or *O. cuniculus*. The results showed that the porcine CR1-like protein was closely related to *T. truncatus* CR1-like protein. In addition, *P. troglodytes*, *Gorilla*, *N. leucogenys*, *M. mulatta*, and *H. sapiens* CR1 or CR1-like protein cluster with relative high bootstrap probability, suggesting that they have a close relationship.

To further investigate the structure of porcine CR1-like protein, we predicted the CR1-like protein structure and functional sites using both Expert Protein Analysis System (ExPAsy) (http://www.expasy.org/) and CBS Prediction Servers software (http://www.cbs.dtu.dk/index.shtml). Like some members of CD superfamily, the deduced porcine CR1-like protein is a hydrophobic single transmembrane protein (**Fig. 1A–C**). The transmembrane region contains 23 amino acids and the cytoplasmic tail has 44 amino acids (**Fig. 1C**). The secondary structure analysis identified a single putative segment of 23 membrane-spanning residues that has strong hydrophobic characteristics. Immediately following this structure are four positively charged residues, Lys–Arg–Lys–Lys (K–R–K–K), a characteristic of many membrane proteins [12]. SignalP software analysis suggested that one signal peptide exists in CR1-like protein and the max cleavage site probability locates between amino acids 40 and 41 (**Fig. 1D**). Its secondary structures were predicted by Self-optimized Prediction Method from Alignment (SOPMA) program (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html). This kind of protein contains large amount of α helix (46.4%) and random coil (40.0%) secondary structures (**Fig. 1E**).

The modification sites in porcine CR1-like protein were also predicted. There are eight potential *N*-glycosylation sites (**Supplementary Fig. S4A**) and no *O*-glycosylation sites are found (**Supplementary Fig. S4B**). Glycosylation is one of ubiquitous post-translational modifications found in all domains of life, and it is important for protein folding, oligomerization, and stability [13]. It is very interesting that predicted porcine CR1-like protein undergoes tense *N*-glycosylation. Whether these modifications are necessary for porcine CR1-like protein function is still unknown. Our previous results demonstrated that porcine erythrocytes are

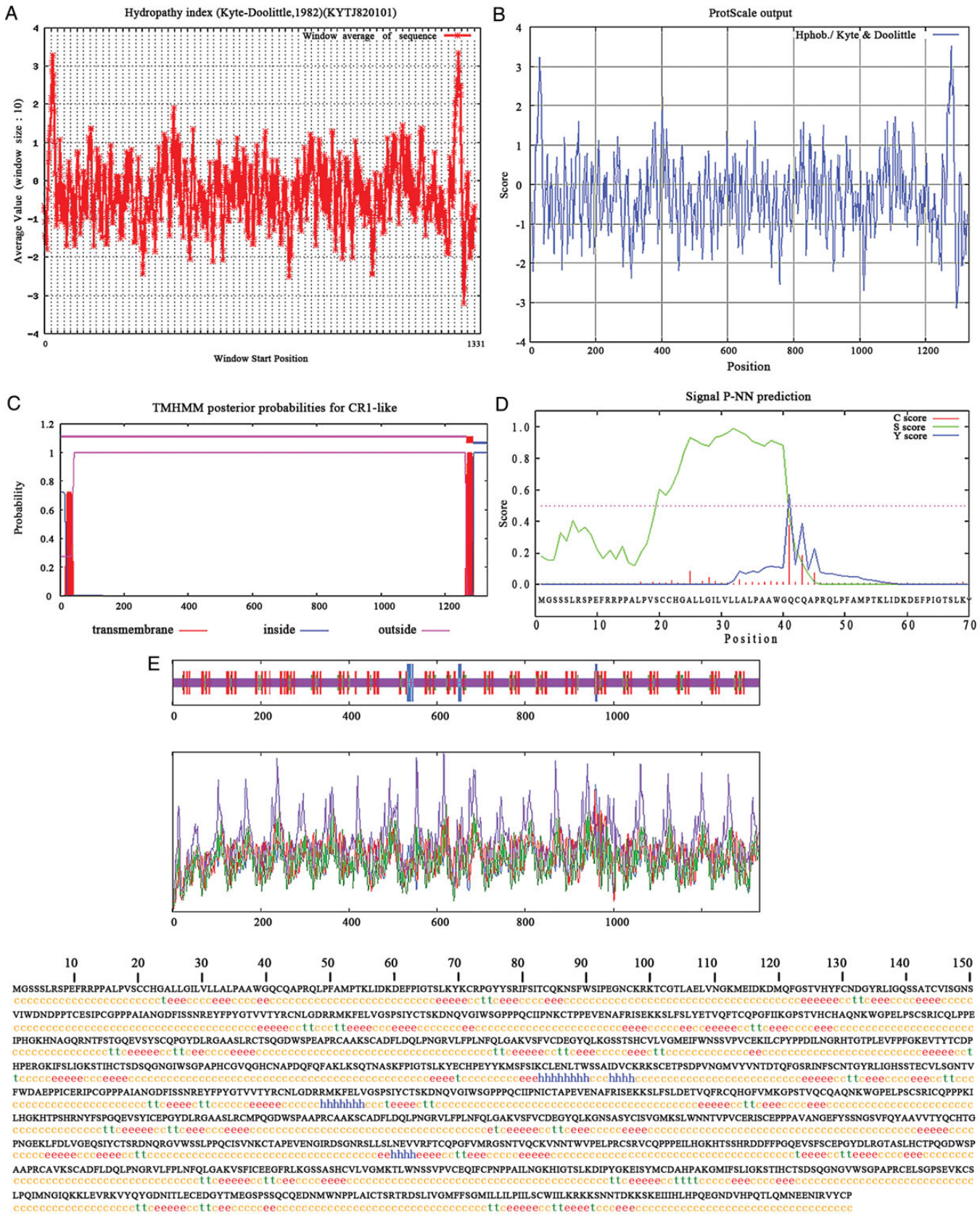

Figure 1. Characterizations of porcine CR1-like protein (A) Analysis of the hydrophobicity of the porcine CR1-like protein using iPSORT. (B) Analysis of the hydrophobicity of the porcine CR1-like protein using ProtScale. (C) Analysis of the transmembrane structure of the porcine CR1-like protein using TMHMM. (D) Analysis of the signal peptide of the porcine CR1-like protein using Signal P 3.0. (E) Analysis of the secondary structure of the porcine CR1-like protein using SOPMA. 'C' represents random coil; 'H' represents alpha helix; 'T' represents beta turn; 'E' represents beta sheet.

able to bind to the opsonized microorganisms. Due to that glycosylation is essential for the immune process [14], it is likely that porcine erythrocytes bind to the opsonized microorganisms through CR1-like protein.

The transcriptional regulation of porcine CR1-like protein has not been addressed so far. Landrace pig genomic DNA was extracted from blood using the QIAamp DNA Blood Midi Kit. CR1-like gene promoter regions were obtained by using seven pairs of primers. We cloned 4974 bp of CR1-like upstream sequence by PCR. The CpG island Searcher was used to search for the CpG island. The porcine CR1-like gene promoter contains relatively abundant CpG sites and island, suggesting that this gene may be regulated by epigenetic modifications (**Supplementary Fig. S5**). MatInspector (Genomatix) was used to analyze the special binding sites in CR1-like gene promoter. Many putative *cis*-regulatory elements were identified in the proximal 5′-flanking region of porcine CR1-like gene, such as FoxP1, GATA2, CHOP, and NFκB. Impressively, a putative NFκB binding site was identified within proximal promoter. NFκB is a conserved transcription factor that regulates the transcription of many inflammatory related genes. It plays a key role in regulating the immune response to infection [15]. If porcine CR1-like protein is involved in the IA, it is reasonable that NFκB binding site exists in the porcine CR1-like gene promoter. Whether NFκB directly regulates porcine CR1-like gene transcription needs to be further explored.

In conclusion, we first cloned the full-length porcine CR1-like cDNA and its promoter region. The bioinformatics analysis confirmed that the porcine CR1-like protein is not conserved, and it possesses the basic structure and characteristics of complement receptor. In our laboratory, further study will focus on designing the primers according to the CR1-like protein sequence, screening specific McAb to obtain CR1-like protein from porcine erythrocytes. We will also focus on whether CR1-like protein directly participates in the IA, and whether the porcine CR1-like protein binds to complement C3b or C4b. If so, the underlying mechanisms will be explored. Further study as to which tissues or cells are involved in the expression of CR1-like gene in porcine will be performed by northern blot and western blot analysis.

Supplementary Data

Supplementary data are available at *ABBS* online.

Funding

This research was supported by the grants from the National Natural Science Foundation of China (No. 31172370) and Research Project Supported by Shanxi Scholarship Council of China (No. 2013060).

References

1. Rowe JA, Moulds JM, Newbold CI and Miller LH. *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. Nature 1997, 388: 292–295.
2. Vik DP and Wong WW. Structure of the gene for the F allele of complement receptor type 1 and sequence of the coding region unique to the S allele. J Immunol 1993, 151: 6214–6224.
3. Ghiran I, Glodek AM, Weaver G, Klickstein LB and Nicholson-Weller A. Ligation of erythrocyte CR1 induces its clustering in complex with scaffolding protein FAP-1. Blood 2008, 112: 3465–3473.
4. Melhorn MI, Brodsky AS, Estanislau J, Khoory JA, Illigens B, Hamachi I and Kurishita Y, *et al*. CR1-mediated ATP release by human red blood cells promotes CR1 clustering and modulates the immune transfer process. J Biol Chem 2013, 288: 31139–31153.
5. Pringle ES, Firth MA, Chattha KS, Hodgins DC and Shewen PE. Expression of complement receptors 1 (CR1/CD35) and 2 (CR2/CD21), and co-signaling molecule CD19 in cattle. Dev Comp Immunol 2012, 38: 487–494.
6. Sun YG, Yin W, Fan XF, Fan KH, Jiang JB and Li HQ. The cytological observation of immune adherence of porcine erythrocyte. Cell Commun Adhes 2012, 19: 79–84.
7. Li H-Q, Lloyd R-J and Wang J-D. Effect of *Astragalus polysaccharides* on erythrocyte immune adherence of chickens inoculated with infectious bursal disease virus. Agric Sci China 2007, 6: 1402–1408.
8. Bork P, Downing AK, Kieffer B and Campbell ID. Structure and distribution of modules in extracellular proteins. Q Rev Biophys 1996, 29: 119–167.
9. Gaboriaud C, Rossi V, Fontecilla-Camps JC and Arlaud GJ. Evolutionary conserved rigid module-domain interactions can be detected at the sequence level: the examples of complement and blood coagulation proteases. J Mol Biol 1998, 282: 459–470.
10. Krych-Goldberg M and Atkinson JP. Structure-function relationships of complement receptor type 1. Immunol Rev 2001, 180: 112–122.
11. Krych M, Clemenza L, Howdeshell D, Hauhart R, Hourcade D and Atkinson JP. Analysis of the functional domains of complement receptor type 1 (C3b/C4b receptor; CD35) by substitution mutagenesis. J Biol Chem 1994, 269: 13273–13278.
12. Klickstein LB, Wong WW, Smith JA, Weis JH, Wilson JG and Fearon DT. Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domains that are composed of the short consensus repeats characteristics of C3/C4 binding proteins. J Exp Med 1987, 165: 1095–1112.
13. Moremen KW, Tiemeyer M and Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol 2012, 13: 448–462.
14. Marth JD and Grewal PK. Mammalian glycosylation in immunity. Nat Rev Immunol 2008, 8: 874–887.
15. Gilmore TD. Introduction to NF-kappaB: players, pathways, perspectives. Oncogene 2006, 25: 6680–6684.