Fish & Shellfish Immunology (2008) **24**, 701–714

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/fsi

Cloning, characterization and expression analysis of SIMP (source of immunodominant MHC-associated peptides) in grass carp *Ctenopharyngodon idella*

Z.Y. Xu [a,b,c], P. Nie [a,b,*], M.X. Chang [a,b], B.J. Sun [a,b]

[a] *State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, Hubei Province, People's Republic of China*
[b] *Laboratory of Fish Diseases, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, Hubei Province, People's Republic of China*
[c] *College of Animal Science and Veterinary Medicine, Huazhong Agricultural University, Wuhan 430070, People's Republic of China*

Received 2 August 2007; revised 25 October 2007; accepted 18 November 2007
Available online 14 April 2008

KEYWORDS
SIMP;
Grass carp;
cDNA;
Gene organization;
Promoter;
Organ

Abstract SIMP (source of immunodominant MHC-associated peptides) plays a key role in N-linked glycosylation with the active site of oligosaccharyltransferase, being the source of MHC-peptides in the MHC I presentation pathway. In the present study, the SIMP gene has been cloned from grass carp *Ctenopharyngodon idella* by rapid amplification of cDNA ends (RACE). The full length of the cDNA sequence is 4384 bp, including a 1117 bp 5′ UTR (untranslated region), a 2418 bp open reading frame, and a 849 bp 3′ UTR. The deduced amino acids of the grass carp SIMP (gcSIMP) are a highly conserved protein with a STT3 domain and 11 transmembrane regions. The gcSIMP spans over more than 24,212 bp in length, containing 16 exons and 15 introns. Most encoding exons, except the first and the 15th, have the same length as those in human and mouse. The gcSIMP promoter contains many putative transcription factor binding sites, such as Oct-1, GCN4, YY1, Sp1, Palpha, TBP, GATA-1, C/EBP beta, and five C/EBP alpha binding sites. The mRNA expression of gcSIMP in different organs was examined by real-time PCR. The gcSIMP was distributed in all the organs examined, with the highest level in brain, followed by the level in the heart, liver, gill, trunk kidney, muscle, head kidney, thymus, and the lowest level in spleen. Furthermore, the recombinant gcSIMP has been constructed successfully and expressed in *Escherichia coli* by using pQE-40 vector, and the polyclonal antibody for rabbit has been successfully obtained, which was verified to be specific. Identification of gcSIMP will help to explore the function in fish innate immunity.

* Corresponding author. State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, Hubei Province, People's Republic of China. Tel.: +86 27 68780736; fax: +86 27 68780123.
E-mail address: pinnie@ihb.ac.cn (P. Nie).

1050-4648/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.fsi.2007.11.019

Introduction

N-linked glycosylation is an essential and evolutionary highly conserved modification of most secretory and membrane proteins occurring in the endoplasmic reticulum (ER) [1—4]. The critical step of this pathway is catalyzed by oligosaccharyltransferase (OST), which mediates the transfer of a preassembled high-mannose oligosaccharide onto asparagine residues of nascent polypeptides. STT3, the most conserved subunit in the OST complex, plays an essential role in substrate recognition and hence for forming part or all of the OST active site [5—8]. Until now, two isoforms of the yeast STT3 protein have been identified in higher eukaryotes, which were designated as STT3A and STT3B [9—11]. The former was initially named conserved transmembrane domain (TMC) [12], while the latter was first identified as a source of immunodominant MHC-associated peptides (SIMP) [10].

Both TMC and SIMP contain the active site that is essential for substrate binding and/or catalytic activity of OST in their C-terminal domain [7,8,13]. However, the mechanism with which they regulate the enzymatic activity of the OST complex is different [11]. In addition to the role in N-linked glycosylation, SIMP also plays roles in the MHC I presentation pathway by being the source of MHC-peptides and displaying a small number of immunopeptidomes for recognition by CD8T cells [14—16], which is not for TMC [10,17—19]. Recently, *N*-glycosylation sites in SIMP were found to make a contribution to SIMP retrotranslocation and degradation, which is one source of MHC I-presented peptides [20,21]. The unique characteristic of SIMP highlights its potential importance in protein modification and immune therapy.

SIMP homologues have been characterized in mammals including human, mouse [10] and avian species, such as chicken. The function of SIMP has been preliminarily studied recently [20,21]. SIMP has first been characterized in grass carp *Ctenopharyngodon idella* by utilizing suppression subtractive hybridization (SSH) in a previous study [22]. Here, the molecular characterization and expression analysis of the grass carp SIMP (gcSIMP) gene, and the putative transcription factor binding sites involved in regulating the gcSIMP gene expression, are firstly reported.

Materials and methods

Cloning the gcSIMP cDNA sequence by RACE-PCR

Total RNA from the liver of healthy grass carp was extracted with Trizol reagent (Invitrogen, USA) according to the manufacturer's instruction. The SMART cDNA was synthesized and amplified using a Clontech SMART PCR cDNA Synthesis Kit (Clontech) by following the supplier's protocol. Based on the obtained sequence of the differential expressed fragment by the suppression subtractive hybridization method (SSH) [22], specific primers were designed, together with the adaptor primer (UPM) for 3′ RACE and 5′ RACE to obtain the full length cDNA sequence of gcSIMP. The full length of gcSIMP was amplified by eight primer pairs, including one primer pair for 3′ RACE, and seven primer pairs for 5′ RACE. During the amplification of the length of the upstream, two degenerated primer pairs, $5F_1/5R_1$ and $5F_2/5R_2$, were used to get part of the conserved sequence, while the other five primer pairs, $5F_{out\ (1,\ 2,\ldots 5)}$ or $5F_{in\ (1,\ 2,\ldots 5)}$, as the first round or second round of 5′ RACE, were used with the adaptor primer (UPM) for amplification, respectively. The position and sequences of all the primers used are listed in Fig. 1. The PCR cycling of 5′ RACE conditions was one cycle of 94 °C for 5 min, nine cycles of 94 °C for 30 s, 68 °C for 30 s and 72 °C for 90 s, 29 cycles of 94 °C for 30 s, 64 °C for 30 s and 72 °C for 90 s, followed by one cycle of 72 °C for 10 min. The 3′ RACE was performed using the following conditions: 94 °C for 5 min, followed by 33 cycles of 94 °C for 5 min, 63 °C for 30 s, 72 °C for 1 min, and then a final elongation step at 72 °C for 10 min. The purified fragments were then cloned into pMD18-T vector (TaKaRa) and transformed into *Escherichia coli* strain DH5α competent cells. The positive clones screened by PCR using the above primers under the same cycle conditions were sequenced using the dideoxy chain-termination method on an automatic DNA sequencer (ABI Applied Biosystems Model 377).

Cloning the gcSIMP genomic sequence and the promoter region

The genomic DNA was purified from the spleen using the Wizard Genomic DNA Purification Kit (Promega). Exon—intron junctions of SIMP were deduced according to the corresponding sequences of other vertebrates from the National Center for Biotechnology Information [23]. Then about 15 primer pairs were designed to obtain the sequence of all the 15 introns respectively, so as to obtain the full length of the corresponding DNA. All the primers used for amplifying the introns of gcSIMP were listed in Table 1.

The promoter region was obtained using a genome walking approach by constructing genomic libraries with a Universal Genome Walker™ Kit (Clontech). Based on the full length cDNA obtained, the gene specific primers (GSP) were designed, together with the adaptor primers for genome walking (AP1 and AP2), and are all listed in Fig. 1. Two specific primer pairs (GSP) were used with AP1 and AP2, respectively. PCR was initially performed with primers $AP1/GSP_{out}$, followed by a second PCR with nested primers $AP2/GSP_{in}$. The PCR cycling conditions were one cycle of 94 °C for 5 min, nine cycles of 94 °C for 30 s, 68 °C for 30 s and 72 °C for 90 s, 29 cycles of 94 °C for 30 s, 64 °C for 30 s and 72 °C for 90 s, followed by one cycle of 72 °C for 10 min. The gel-purified PCR products were subcloned into the pMD18-T vector (TaKaRa) and sequenced.

Sequence analysis

Protein prediction was performed using software at the ExPASy Molecular Biology Server (http://expasy.pku.edu.cn). The putative transmembrane sequence was identified by SOSUI [24]. The model of the proposed SIMP topology (Fig. 2) is based on SOSUI and TOPO2 (http://www.sacs.ucsf.edu/TOPO2/) prediction methods [20]. Multiple alignments were generated at the web site (http://searchlauncher.