HELMINTHOLOGIA, 41, 1: 15 – 19, 2004

Molecular cloning and characterization of a novel gene encoding tetraspanin of *Schistosoma japonicum*

Z. WU, X. YU, J. XU, Y. ZHENG, A. CAO

Department of Parasitology, Zhongshen Medical College, Sun Yet-sen University, Guangzhou 510089, China,
E-mail: *zdwu62@263.net*

Summary

In this study, a novel tetraspanin-like gene product (Sj04/TM4) was identified. An EST named JAYN004.GYL (Gen Bank accession number AW231251) matched the Sj25 gene of *Schistosoma japonicum* with 97 % (539/554) identity. The longest cDNA insert (1079 bp) was obtained from the recombinant clone JAYN004.GY by PCR. Sequence analysis showed that it contains a complete open reading frame (ORF) coding for a protein of 224 amino acids with a predicted molecular mass of 25,622.58 Dalton. Alignment analysis with the entire ORF revealed that it had 93.3 % identity in 224 aa overlap % identity to Sj25/TM from *S. japonicum*. It also exhibited 31.4 % in 220 aa overlap and 21.1 % in 209 aa overlap, respectively to tetraspanin TE736 from *S. japonicum* and Sm23 antigen from *S. mansoni*, whose are belong to the members of tetraspanins. Analysis of amino acid sequence showed that this gene product contains all the conserved subdomains found in other members of tetraspanin superfamily. Hydropathy analysis of the deduced amino acid sequence revealed a profile very similar to those of other tetraspanins containing 4 transmembrane domains. Taken together, Sj04/TM4 is a novel gene encoding tetraspanin-like protein, or belongs to a subtype of Sj25/TM4 tetraspanin superfamily.

Key words: tetraspanin; EST; molecular cloning; *Schistosoma japonicum*

Introduction

Tetraspanins, or the transmembrane 4 superfamily (TM4 SF), are a group of hydrophobic proteins with four transmembrane domains and two extracellular (EC) loops, both with conserved residues. Some tetraspanins are cell specific and others are very ubiquitous. Tetraspanins interact with very different types of proteins such as integrins, membrane receptors, as well as intracellular signaling molecules. Genetic evidence points to critical roles for tetraspanins on fertilization, adhesion and invasion, cell activation, proliferation, differentiation, motility, and signal transduction (Yunta *et al.*, 2003). Tetraspanin proteins are characterized by containing four transmembrane Δ-helices, which have several hydrophilic redidues, and two EC (Hemler *et al.*, 2001). Tetraspanins are abundantly expressed transmembrane proteins (25 – 50 kDa), with at least 30 distinct family members in mammals, 37 members in *Drosophila* and 20 in *Caenorhabditis elegans* (Maecker *et al.*, 1997; Stipp *et al.*, 2003). In 1997, the first member of TM4SF from *S. mansoni*, termed Sm23 was reported by Wright and his colleagues (Wright *et al.*, 1990; Lee *et al.*, 1995). Some others such as Sj25/TM4 (Fan *et al.*, 1997), TE736 (Fan *et al.*, 1998a) and SmTSP-1 (Smyth *et al.*, 2003) were isolated from *S. japonicum* and *S. mansoni*. The members of the tetraspanin superfamily have been conserved during evolution after diverging from a common ancestral gene. Some TM4SF, including human CD81, CD82, CD9 and CD63 are expressed in virtually all tissues, whereas the expression of others such as human CD37 and CD53 are highly restricted in B cells and lymphoid/myeloid cells, respectively.

In the present study, we report the cloning, sequencing, and characterization of a cDNA clone (JAYN004.GYL) encoding a novel TM4SF from *S. japonicum*, designated Sj04/TM4. Sj04/TM4 is a 25,622.58 Dalton surface protein composed of four transmembrane and two extracellular domains and an overall topology shared by the superfamily of proteins known as tetraspanins, or the TM4SF.

Materials and Methods

Schistosome cDNA library

The adult worms of *S. japonicum* (central China isolate) were perfused and collected from the mesenteries of expe-

rimentally infected rabbits. Double-stranded cDNA synthesized with the mRNA isolated from adult worms, was inserted into the bacteriophage 0 gt11 Sfi-Not arms between EcoR I and Not I sites of the LacZ gene. The cDNA library was contructed by Chen *et al.* (1997) at Nanjing Medical University, Nanjing, Jiangsu.

Identification of expressed sequence tags

Clones from the cDNA library were randomly selected and used in single pass automated DNA sequencing to produce expressed sequence togs (ESTs) as described (Wu *et al.*, 2000; Bian *et al.*, 2002). Identification of ESTs obtained from one cDNA strand was performed by comparison with all DNA and protein sequences deposited in the EMBL-parasites database (http://www2.ebi.ac.uk/blast2/parasites.html) by using Basic local alignment Search tool (BLAST n) program. Sequences that did not match sequences in the nucleotide database were further analyzed by searching for similarities at protein level by using BLASTx or at the nucleotide level by using BLASTn on the BLAST network service at NCBI (http://www.ncbi.nlm.nih.gov/blast). Sequence in which the alignments scored 200 or more for BLASTn and 100 or more for BLASTx were selected for further analysis, and the alignments were assigned a putative identification (Fan *et al.*, 1998b).

Cloning and analysis of full-length cDNA

According to the homology analysis, the interesting ESTs were selected to clone the full-length cDNA. Among them, one EST named JAYN004.GYL matched the Sj25 gene of *S. japonicum*. The cDNA insert of the clone JAYN004.GY was obtained by PCR. The primers were designed based on the sequence of the bacteriophage 0 gt11 and were as follows:

Forward primer

5'-GGTGGCGACGACTCCTGGAGCCCG

Reverse primer

5'-TTGACACCAGGCCAACTGGTAATG

The cDNA sequence was subcloned into pGEX-4T-1 vector and sequenced as described before. Sequence and motif analysis including functional sites and motifs on target proteins were located using Proteomics tools such as ProtScan, InterProScan and ScanProsite etc at http://www.expasy.org/tools/.

Results and Discussion

A new gene encoding tetraspanin from schistosomes

During our analysis of 531 ESTs from *S. japonicum*, we observed that an EST named JAYN004.GYL (GenBank accession nos. AW231251) matched the Sj25 gene of *S. japonicum* (EMBL accession nos.U77941), the identities was 97 % (539/554). The complete cDNA insert of the clone (JAYN004.GY) is 1079bp long. The cDNA was translated into six possible frames, and the length of the longest open reading frame (ORF) was 675 bp, encoding a predicted 224-aa (amino acids, aa) protein. The deduced polypeptide had a predicated molecular mass 25622.58 Dalton and a theoretical isoelectric point of 8.92.

During the analysis of 531 ESTs from *S. japonicum*, were found that one EST clone named JAYN004.GYL (Genbank accession number AW231251) matched the Sj25 gene of *S. japonicum* (EMBL accession U77941) with an identity of 97 % (539/554). The complete cDNA insert of the clone (JAYN004.GY) is 1079bp long. The cDNA was translated into six possible frames, and the length of the longest ORF was 675bp, coding for a protein of 224 amino acids with a predicted molecular mass of 25 622.58 Dalton and a theoretical isoelectric point of 8.92.

Alignment analysis using the 224aa of the entire ORF revealed a highly significant similarity with *Sj*25/TM4 (Genbank accession nos. P91799), which identity between the sequence of JAYN004.GY and Sj25/TM4 is 93.3 % in 224 aa overlap (209/224). Alignment analysis also revealed lower similarity with other members of the tetraspanin family including tetraspanin TE736 (Genbank accession nos. AAC69992) from *S. japonicum* and Sm23 antigen (Genbank accession nos. P27591) from *S. mansoni*, the identity is 31.4 % in 220 aa overlap and 21.1 % in 209 aa overlap, respectively. These results suggested that the clone JAYN 004.GY encodes a novel tetraspanin-like Sj25/TM4 protein, designated Sj04/TM4.

Transmembrane domains and other functional sites of Sj04/TM4

Sj04/TM4 consists of four hydrophobic, putative transmembrane regions, consistent with the characteristics of the tetraspanin superfamily. The positions of the four transmembrane domains (TM1-4), characteristic of tetraspanins are at aa13 – 35, aa53 – 75, aa82 – 104, and aa186 – 208, respectively. Three of the transmembrane domains (TM1-3) are clustered at the NH_2 –terminus, followed by a single hydrophilic domain, with the fourth hydrophobic domain (TM4) located at the COOH-terminus (Fig. 3).

Five potential functional sites are present in Fig. 1 as follows: they are N-glycosylation site, Protein kinase C phosphorylation site (SsK, TcR, SeK), Casein kinase II phosphorylation site (TiiD), Tyrosine kinase phosphorylation site (RyyvDiliY), N-myristoylation site (GvrpSI, GLmkGG, GQvlND). These motifs are present in other members of tetraspaninin superfamily such as TE736 from *S. japonicum* (Fan *et al.*, 1998).

Hydrophobic and hydrophilic profile of Sj04/TM4

The analysis of hydrophobic and hydrophilic regions of the predicted amino acid sequence of Sj04/TM4 exhibited the four hydrophobic domains, which are in the same positions as TM1, TM2, TM3 and TM4, respectively. The plot shows the four potential transmembrane domains TM1 through TM4 are hydrophobic domains. The large extracellular loop of the tetraspanins, which locates between TM3 and TM4, exhibits a hydrophilic domain (Figs. 3,4). Most of the variable amino acids, comparing Sj04/TM4 and Sj25/TM4, are also present in the extracellular domain (Fig. 2). The characteristic region is probably involved in antigenic epitopes and polymorphism of Sj25-like tetra-

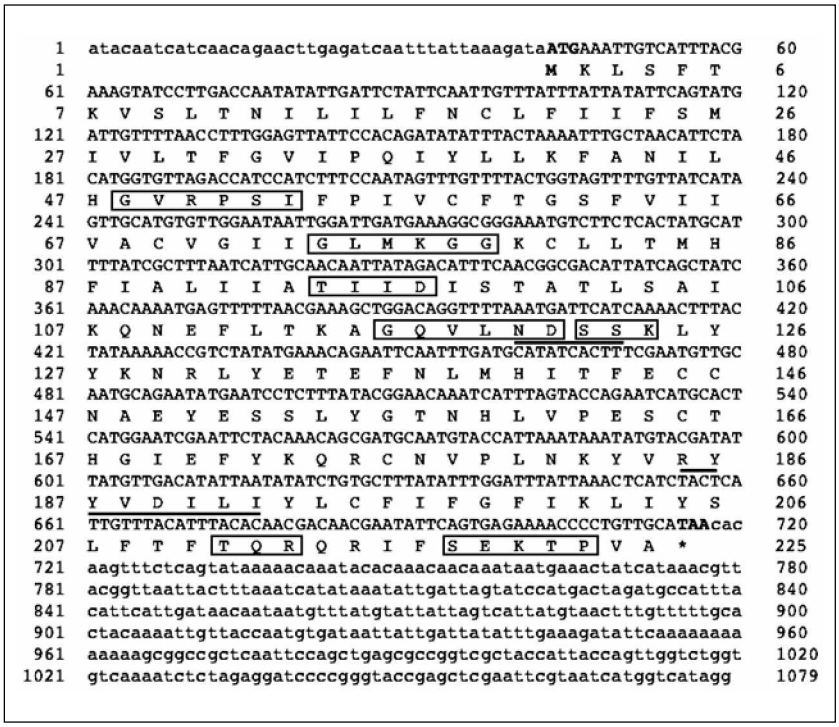

Fig. 1. Nucleotide and deduced amino acid sequence of cDNA clone JAYN004 (GCG Program). Predicted start and stop codons are in bold. Putative functional sites are indicated by underlining/italics and box: N-glycosylation site (aa120 – 123), Protein kinase C phosphorylation site(aa122 – 124 SsK, aa211 – 213 TcR, aa 218 – 220 SeK), Casein kinase II phosphorylation site(aa94 – 97 TiiD), Tyrosine kinase phosphorylation site(aa185 – 193 RyyvDiliY), N-myristoylation site (aa48 – 53 GvrpSI, aa74 – 79 GLmkGG, aa116 – 121 GQvlND)

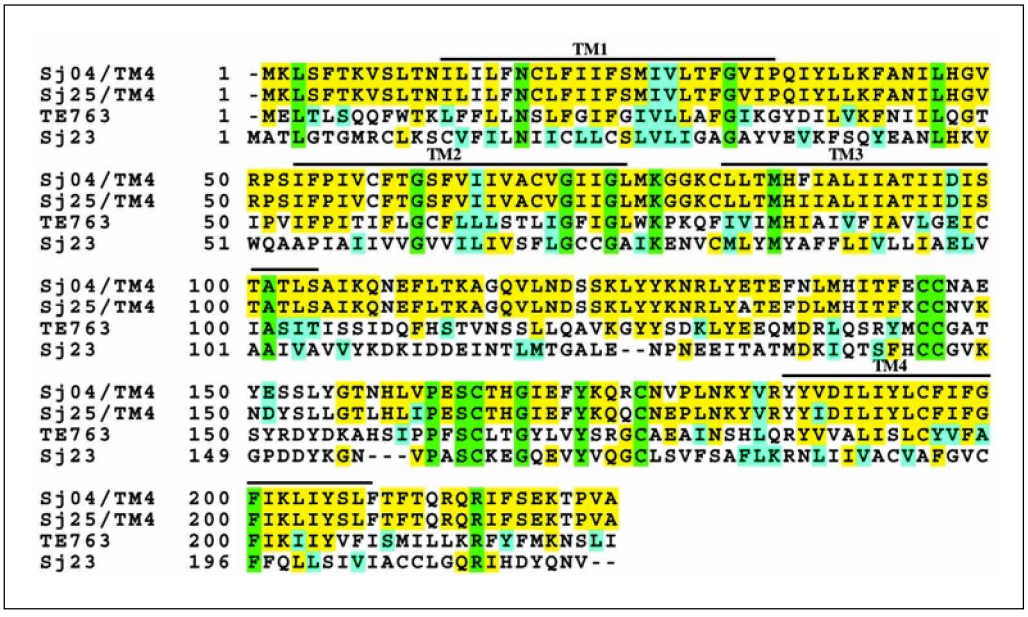

Fig 2. ClustalW (Thompson *et al.*, 1994) alignment of tetraspanin protein amino acid sequences from *S. japonicum* (Sj04/TM4, AW231251; Sj25/TM4, P91799; TE736, AAC69992; and Sj23, P27591). Completely conserved residues are shown in green, identical residues in yellow, similar residues in cyan, and different residues in white. Amino acid residues absent from other sequences are denoted by dashes. There are 15 variable amino acids from aa87 to aa188 comprising both sequences (aa87 FOI, aa133 EOA, aa137 NOD, aa144 EOK, aa148 AOV, aa149 EOK, aa150 YON, aa151 EOD, aa152 SOY, aa155 YOL, aa158 NOL, aa161 VOI, aa175 ROQ, aa178 VOE, aa188 VOI). TM1-4 are indicated by solid lines above the alignment

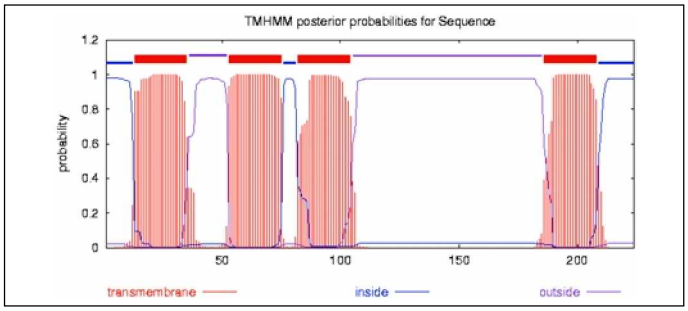

Fig. 3. Analysis of the amino acid sequence of Sj04/TM4 using TMHMM program (http://www.expasy.org/tools/), predicating four-transmembrane hydrophobic domains. Shaded bars indicate the positions of transmembrane domains

spanin family. Since the Sj23, a homolog of Sj25, is an important target for a vaccine against schistosomosis, it is suggested that Sj04/TM is a new potential target for future research.

Conclusions

The full-length cDNA encoding a member of tetraspanin superfamily termed Sj04/TM4 was isolated and cloned by Expressed Sequence Tags strategy. Based on the closed identity between Sj04/TM4 and Sj25/TM4, as well as other tetraspanins, we suggest that Sj04/TM4 is a novel tetraspanin-like protein, or a subtype of Sj25/TM4 tetraspanin superfamily. Therefore, there are at least four distinct tetraspanins in *S. japonicum*, Sj23, Sj25, TE736 and Sj04/TM4 characterized so far. The expression of recombinant Sj04/TM4 and its immunogenisity are under investigation.

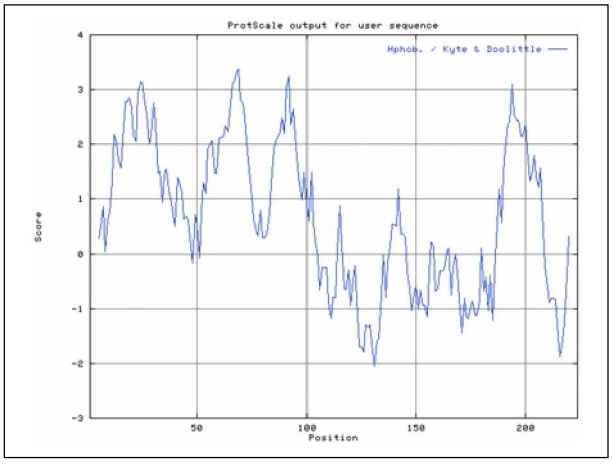

Fig. 4. The analysis of hydrophobic and hydrophilic regions of the predicted amino acid sequence of Sj04/TM4. Hydrophobic residues are positive. The plot was obtained using the ProtScale program (http://www.expasy.org/tools/)

Acknowledgement

We would like to thank Dr. M. Fung (Hong Kong Chinese University) and Dr. S. Luo (University of Illinois) for critically reading the manuscript.
The project was supported by National Natural Science Foundation of China (No. 30070683).

References

BIAN, G. W., YU, X. B., WU, Z. D. (2002): Generation and analysis of 113 adult *stage Schistosoma japonicum* (Chinese strain) expressed sequence tags. *Chin. Med. J.*, 115: 1517 – 1520

CHEN, S. Z., WU, H. W., ZHANG, S. S., WU, Y. Q. (1997): A directional cloning gene expression library of *Schistosoma japonicum*. *Chin. J. Zoonoses*, 13: 23 – 25

FAN, J., HOOKER, C. W., MCMANUS, D. P., BRINDLEY, P. J. (1997). A new member of the transmembrane 4 superfamily (TM4SF) of proteins from schistosomes, expressed by larval and adult *Schistosoma japonicum*. *Biochim. Biophys. Acta*, 1329: 18 – 25

FAN, J., BRINDLEY, P. J. (1998a): Characterization of cDNAs encoding a new family of tetraspanins from schistosomes- the Sj25 family. *Gene*, 219: 1 – 8

FAN, J., MINCHELLA, D. J., DAY, S. R., MCMANUS, D. P., TIU, W. U., BRINDLEY, P. J. (1998b): Generation, identification, and evaluation of expressed sequence tags from different developmental stages of the Asian blood fluke *Schistosoma japonicum*. *Biochem. Biophys. Res. Comm.*, 252: 348 – 56

HEMLER, M. E. (2001): Specific tetraspanin functions. *J. Cell. Biol.*, 155: 1103-1107

LEE, K. W., SHALABY, K. A., MEDHAT, A. M., SHI, H., YANG, Q., KARIM, A. M., LOVERDE, P. T. (1995): *Schistosoma mansoni*: characterization of the gene encoding Sm23, an integral membrane protein. *J. Exp. Parasitol*, 80: 155 – 158

MAECKER, H. T., TODD, S. C., LEVY, S. (1997): The tetraspanin superfamily: molecular facilitators. *FASEB J.*, 11: 428 – 442

SMYTH, D., MCMANUS, D. P., SMOUT, M. J., LAHA, T., ZHANG, W., LOUKAS, A. (2003): Isolation of cDNAs encoding secreted and transmembrane proteins from *Schistosoma mansoni* by a signal sequence trap method. *Infect. Immun.*, 71: 2548 – 2554

STIPP, C. S., KOLESNIKOVA, T. V., HEMLER, M. E. (2003):

Functional domains in tetraspanin proteins. *Trends Biochem. Sci.*, 28: 106 – 112
THOMPSON, J. D., STAHELIN, T., GIBSON, T. J. (1994): CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nuc. Acid. Res.*, 22: 4673 – 4680
WU, Z. D., YU, X. B., ZHENG, Y. N., LI, Y., ZHOU, J. M. (2000): Isolation and identification of expressed *Schistosoma japonicum* Chinese strain genes from a directional cDNA library. *Chin J. Zoonoses*, 16: 3 – 6
WRIGHT, M. D., HENKLE, K. J., MITCHELL, G. F. (1990): An immunogenic Mr 23,000 integral membrane protein of *Schistosoma mansoni* worms that closely resembles a human tumor-associated antigen. *J. Immunol.*, 144: 3195 – 3200
YUNTA, M., LAZO, P. A. (2003): Tetraspanin proteins as organizers of membrane microdomains and sigalling complexes. *Cell. Signal.*, 15: 559 – 564

RECEIVED APRIL 23, 2003

ACCEPTED FEBRUARY 2, 2004