SPANDIDOS PUBLICATIONS

In silico study of a novel gene evolved from an ancestral *SVIP* gene and highly expressed in the adult mouse testes

MOMOKO YOSHIDA[1,2], AKIFUMI YAMASHITA[2], YOSHIKO IDOJI[1], SEIJI NISHIGUCHI[1], KAZUNORI SHIMADA[1], TERUO YASUNAGA[2] and HIROMICHI YAMANISHI[1]

[1]Hirakata Ryoikuen, 2-1-1 Tsudahigashi, Hirakata, Osaka 573-0122; [2]Department of Genome Informatics, Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Received February 29, 2008; Accepted April 23, 2008

DOI: 10.3892/ijmm_00000001

Abstract. We found that a cDNA clone isolated from a mouse testis cDNA library, 1700015G11 (Mmu_15G11), corresponded to the most highly expressed testis-specific mRNA in the adult mouse. Although the Mmu_15G11 cDNA is predicted to encode a small protein consisting of 67 amino acid residues, it has not yet been functionally annotated and has been designated as an unclassifiable clone. Since the Mmu_15G11 protein possibly has a pivotal role in spermatogenesis, we initiated an *in silico* study of this clone, and revealed that an ancestral gene of *15G11* genes evolved from an ancestral gene for mammalian small valosin-containing protein-interacting protein (*SVIP*) genes by gene duplication. Although SVIP protein reportedly participates in endoplasmic reticulum-related protein degradation, 15G11 protein is predicted to be a nuclear protein and possibly participates in the interaction between proteins and nuclear DNA.

Introduction

We initiated a study of testis-specific genes using the Novartis expression database (1) because it contains a dataset of the expression levels of 36,182 mRNAs from 61 different mouse tissues by two independent experiments. We accessed the Novartis expression database, Mouse GNF1M (1), and tried to isolate cDNA clones corresponding to highly expressed testis-specific genes, and found that one of the cDNA clones, 1700015G11, corresponded to the most highly expressed testis-specific gene. Although 1700015G11 cDNA is predicted to encode a small protein consisting of 67 amino acid residues, it has so far been designated as an unclassifiable clone (2,3). We believe that the proteins encoded by the most highly expressed mRNAs in the adult mouse testes should have pivotal roles in spermatogenesis, and initiated an *in silico* study of this cDNA clone (1700015G11, referred to as Mmu_15G11 herein).

Our current study revealed that an ancestral gene for the mammalian *15G11* genes evolved from an ancestral gene for the mammalian small valosin-containing protein-interacting protein (*SVIP*) genes by gene duplication. The rat SVIP protein is anchored to the microsomal membrane via myristoylation and co-fractionates together with several other proteins including valosin-containing protein/97-kDa protein (VCP/p97; p97 is a synonym of VCP) to the endoplasmic reticulum (ER) (4,5). The rat SVIP has recently been reported to be an endogenous inhibitor of ER-associated degradation (ERAD) of several proteins (6). Although mammalian *15G11* genes have evolved from an ancestral gene for mammalian *SVIP* genes, we predict that the functions of mammalian 15G11 proteins are quite different from that of the mammalian SVIP proteins, i.e., 15G11 proteins are nuclear proteins and possibly participate in the interaction between proteins and nuclear DNA.

Materials and methods

Isolation of cDNA clones corresponding to the most highly expressed testis-specific mRNAs. We selected cDNA clones corresponding to the most highly expressed testis-specific mRNAs from the expression database 'Mouse GNF1M (MAS5-condensed)' (http://wombat.gnf.org/index.html) (1). Log-transformed experimental data were averaged and normalized (7). The highest standard deviation (SD) for the microarray data from the testes was around mean +7.3 SD, and 20 of the 36,182 expression values exceeded mean +7.0 SD (data not shown). Although we found that two clones, 4922502D21 and 1700015G11, showed mean +7.3 SD, we selected 1700015G11 (Mmu_15G11) for this study because

Correspondence to: Dr Hiromichi Yamanishi, Hirakata Ryoikuen, 2-1-1 Tsudahigashi, Hirakata, Osaka 573-0122, Japan
E-mail: hirochan@hirakataryoiku-med.or.jp

Abbreviations: ER, endoplasmic reticulum; ERAD, ER-associated degradation; EST, expressed sequence tag; NLS, nuclear localization signal; NLS_BP, bipartite_NLS; SD, standard deviation; SVIP, small valosin-containing protein-interacting protein; VCP/p97, valosin-containing protein/97-kDa protein (p97 is a synonym of VCP); VIM, VCP/p97-interacting motif

Key words: gene duplication, MOTIF search, multiple alignments, nuclear localization signal, phylogenetic tree

Table I. List of 15G11 and SVIP sequence identifiers used in this study.

Name	Organism	Nucleotide	Protein	UniGene
Hsa_15G11	*Homo sapiens*	XM_932981	XP_938074	Hs.555029
Mmu_15G11	*Mus musculus*	XM_895470	XP_900563	Mm.279725
Rno_15G11	*Rattus norvegicus*	XM_579921	XP_579921	Rn.52198
Cfa_15G11	*Canis lupus familiaris*	XM_860537	XP_865630	Cfa.7281
Bta_15G11	*Bos taurus*	EH124756[a]	See note[b]	Bt.90924
Cfa_15&SV	*Canis lupus familiaris*	XM_534092	XP_534092	Cfa.7281
Hsa_SVIP	*Homo sapiens*	NM_148893	NP_683691	Hs.349096
Mmu_SVIP	*Mus musculus*	AK133737	BAE21812	Mm.386823
Rno_SVIP	*Rattus norvegicus*	XM_574455	XP_574455	Rn.77399
Cfa_SVIP	*Canis lupus familiaris*	DN426689[a]	See note[b]	Cfa.7281
Bta_SVIP	*Bos taurus*	EH140382[a]	See note[b]	Bt.97558
Tvu_SVIP	*Trichosurus vulpecula*	EG609857[a]	See note[b]	Tvu.8757
Gga_SVIP	*Gallus gallus*	XM_001234337	XP_001234338	Gga.44586
Xla_SVIP	*Xenopus laevis*	EE314924[a]	See note[b]	Xl.23091
Dre_SVIP	*Danio rerio*	XM_001334239	XP_001334275	Dr.85238

[a]Indicates EST sequence. [b]The coding region of the EST sequence was translated.

it has so far been annotated as 'unclassifiable', whereas 4922502D21 has already been annotated as coding for a C-type lectin domain containing protein (2,3).

Nomenclature. The nomenclatures of proteins and genes are shown in uppercase, and gene names are also italicised.

Collecting orthologues of mouse 15G11 and SVIP proteins. Using either the Mmu_15G11 or Mmu_SVIP amino acid sequences as queries, we performed homology searches with the blastp program against a set of databases (all non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects, 5,789,131 sequences; 1,995,680,057 total letters), and with the tblastn program against a different set of databases (GenBank non-mouse and non-human EST entries, 35,344,825 sequences; 20,271,993,704 total letters) (8).

Structures and locations of exon-intron junctions. To determine the exon-intron junctions, the collected cDNA and/or EST sequences were aligned to the corresponding genomic sequences using the BLAT DNA analysis program (9).

Multiple alignments and phylogenetic tree. The collected amino acid sequences of the cDNAs and translations of the coding regions of the ESTs were aligned. Multiple alignments were performed with CLUSTAL W (10). Phylogenetic and molecular evolutionary analyses were conducted by the neighbor-joining (NJ) method (11) in MEGA version 4.0 (12).

Searches with protein query sequences against 'Motif Libraries', and prediction of functions carried out by 15G11 proteins. The functions of 15G11 proteins were predicted by searching several motif databases, such as PROSITE, BLOCKS, ProDom, PRINTS, and Pfam, using the amino acid sequences of 15G11 or SVIP proteins as queries for MOTIF search in GenomeNet (13,14).

Estimating the levels and distributions of mRNAs. The expression profiles of Mmu_15G11 and Mmu_SVIP mRNAs were confirmed by accessing the dataset in the Mouse GeneAtlas GNF1M, MAS5 (http://wombat.gnf.org/SymAtlas/) (1).

Results

Isolation of orthologous proteins for mouse 15G11. To elucidate the function of Mmu_15G11, we performed blastp homology searches, and found several Mmu_15G11 orthologues from human, rat, and dog (Table I). All these proteins, except for the one derived from dog, are small proteins consisting of fewer than 70 amino acid residues, and so far they have no functional annotations. Interestingly, we found that the homologue from dog, named Cfa_15&SV (Table I), consisted of 124 amino acid residues and contained two different domains; one domain was highly homologous to Mmu_15G11, and the other domain to a protein known as mouse SVIP (Fig. 1A, Mmu_SVIP). Subsequently, we found that there is a weak but significant homology between the Mmu_15G11 and Mmu_SVIP proteins (Fig. 1B).

Evolution of the mammalian 15G11 genes. To further confirm the above, we performed homology searches using Mmu_15G11 or Mmu_SVIP as queries and screened the NCBI database. After running the blastp program, we

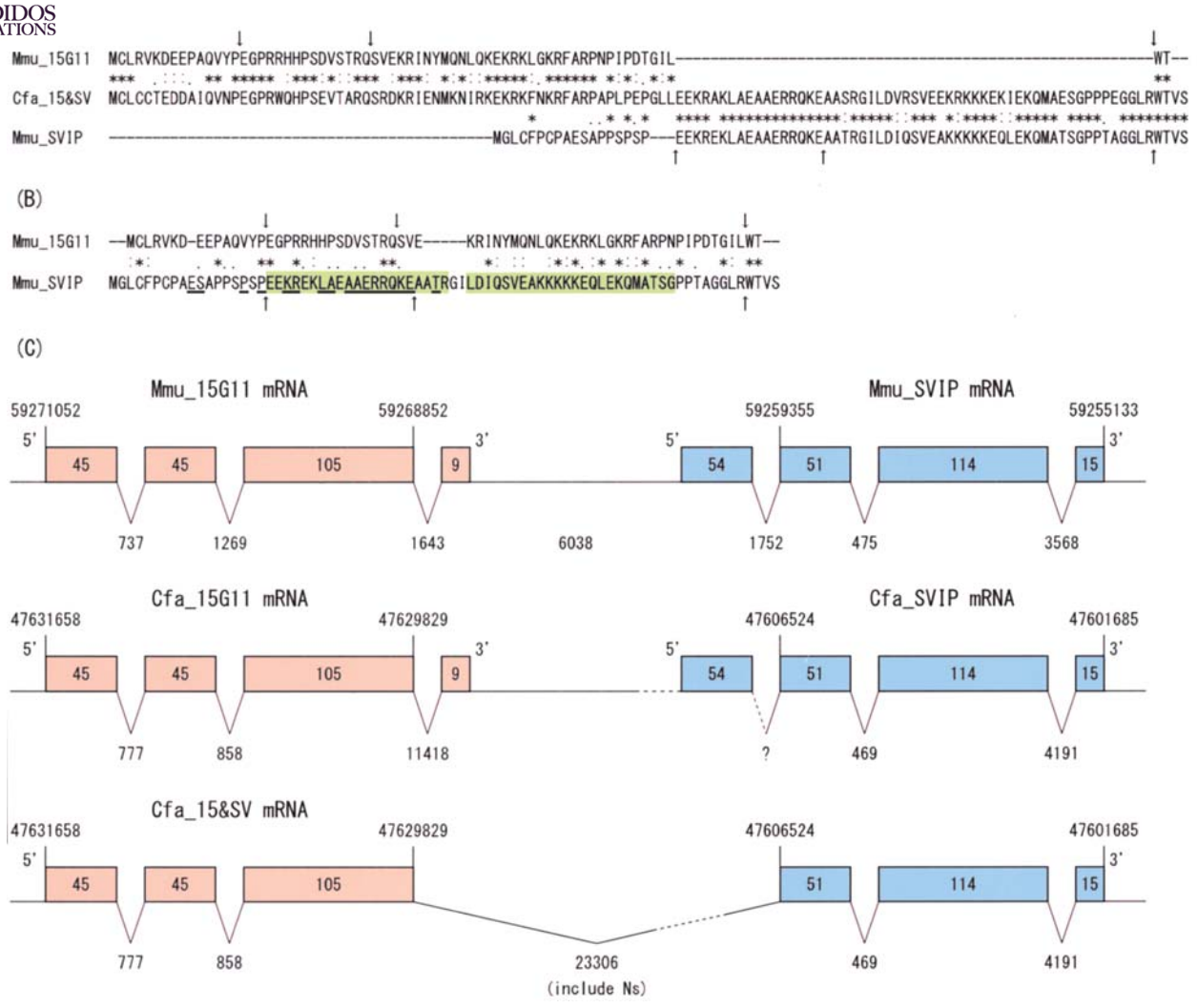

Figure 1. Amino acid sequence alignments and transcription of the *15G11* and *SVIP* genes. Amino acid sequences of Mmu_15G11, Cfa_15&SV and Mmu_SVIP (A), and Mmu_15G11 and Mmu_SVIP (B), were aligned as described in Materials and methods. For the conservation line output in the CLUSTAL W alignments, three characters are used: '*' indicates positions that have a single, fully conserved residue; ':' and '.' indicate positions that contain substitutions with highly and with weakly similar amino acid residues, respectively. Arrows indicate the sites of introns. In panel (B) Mmu_SVIP line, the green-colored regions show the two coiled-coil regions, and the underlined amino acid residues correspond to the conserved amino acid residues of the VIM-like sequence, which was identified within the N-terminal 39 amino acid regions of human, mouse, and rat SVIP (5). (C) Transcription of *Mmu_15G11*, *Mmu_SVIP*, *Cfa_15G11*, and *Cfa_SVIP* genes, and an alternative splicing observed in one of the transcripts of *Cfa_15G11* and *Cfa_SVIP* genes. The *Mmu_15G11* and *Mmu_SVIP* genes mapped on mouse chromosome 7 and the *Cfa_15G11* and *Cfa_SVIP* genes on dog chromosome 21. The coding exons of the *15G11* genes are indicated by brown boxes, and those of the *SVIP* by blue boxes. Spliced introns and their sizes are indicated by v lines with the numbers of nucleotides, and the numbers within the boxes indicate the sizes of coding exons in nucleotides. Since several parts of the dog genome including the regions covering the 5'- and 3'-ends of the 1st coding exon of *SVIP* gene have not yet been completely sequenced and contained N sequences, the exact position of the 5'-end of the 1st coding exon of the dog *SVIP* gene has not yet been determined. Dotted lines indicate that the region contains N-sequenced regions. The end of the 5'-untranslated region is indicated by 5', and 3' indicates the start of the 3'-untranslated region. The large numbers shown over or close to the 5' and/or 3' marks correspond to the sequenced genome nucleotide numbers. The other names and abbreviations are summarized in Table I.

collected four 15G11 and five SVIP orthologues (Table I). After running the tblastn program, we collected one bovine EST as encoding a 15G11 orthologue, and one EST each from dog, bovine, opossum, and frog encoding SVIP orthologues (Table I). Notably, five Mmu_15G11 orthologues were isolated from mammals, but none from the other lower organisms, whereas Mmu_SVIP orthologues were isolated not only from mammalian sources but also from chicken, frog, and fish (Table I). We constructed a phylogenetic tree using all these 15G11 and SVIP orthologues (Fig. 2).

The phylogenetic tree constructed from all these protein sequences revealed that an ancestral gene for mammalian *15G11* genes evolved from an ancestral gene for mammalian *SVIP* genes by gene duplication (see Fig. 2, Y-shaped bold line). After the duplication, the ancestral gene for mammalian *SVIP* genes seemed to diverge at a relatively stable pace, whereas an ancestral gene for the *15G11* genes seemed to initiate its divergence after a phase of rapid evolution, and the divergence itself progressed significantly faster than that of the mammalian *SVIP* genes (Fig. 2).

Structures of the mammalian 15G11 and SVIP genes. We found that *15G11* and *SVIP* genes are arranged side by side in mouse, dog, human, and rat chromosomes. The distance between the 3'-end of the 4th coding exon of *Mmu_15G11* and the 5'-end of the 1st coding exon of *Mmu_SVIP* genes

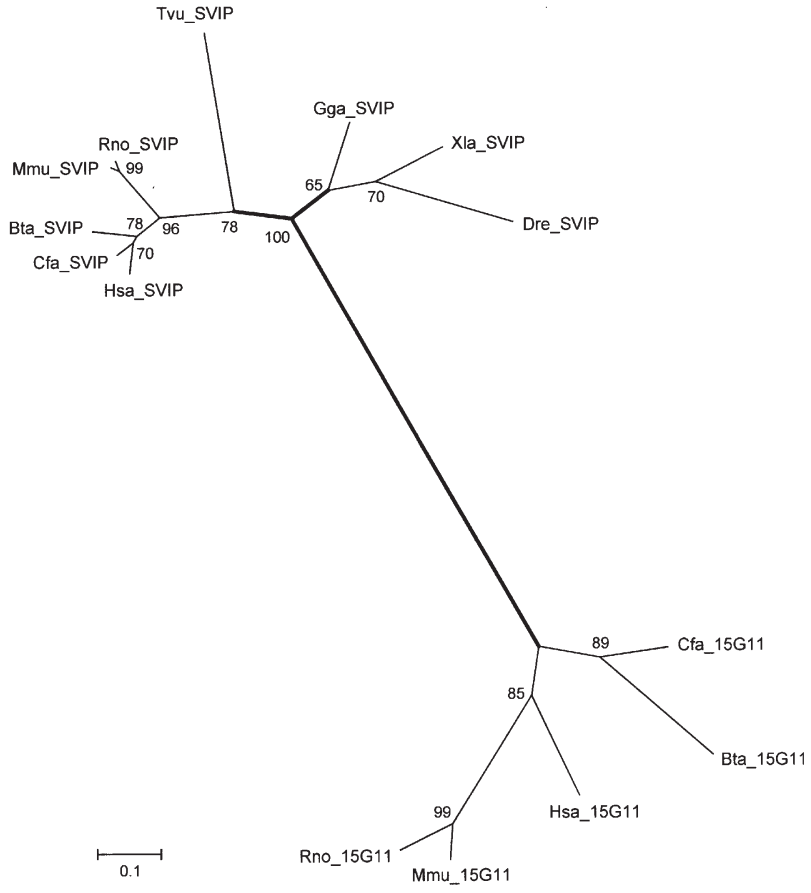

Figure 2. A phylogenetic tree of 15G11 and SVIP proteins. This tree was constructed by aligning the amino acid sequences of the 15G11 and SVIP proteins listed in Table I. Phylogenetic and molecular evolutionary analyses were conducted using MEGA software version 4 (MEGA4) as described in Materials and methods. Bootstrap values obtained from 1,000 replicates are shown at the corresponding branches by the 1/10 values. The Y-shaped bold line indicates the evolution of ancestral genes for the mammalian SVIP and 15G11 proteins. The other abbreviations are as described in Table I.

was 6,038 bp (Fig. 1C), and the corresponding distances were 17,761 bp in human and 10,397 bp in rat chromosomes (data not shown). The distributions of introns revealed that Mmu_15G11, Mmu_SVIP, Cfa_15G11, and Cfa_SVIP proteins were all encoded by four exons, which supported the assertion that they are the products of gene duplication (Fig. 1C).

As described earlier, we found that one of the dog cDNA clones, Cfa_15&SV, encodes a 15G11-SVIP-fused protein (Fig. 1A). We collected together fourteen 15G11 and ten SVIP cDNAs or ESTs from dog and found that only one of them corresponded to the Cfa_15&SV mRNA. This result suggests that Cfa_15&SV may be an artifact produced during the course of cDNA cloning. However, since the 5'- and 3'-ends of the 3rd intron of the *Cfa_15&SV* gene corresponded exactly to the 5'-end of the 3rd intron of the *Cfa_15G11* gene and the 3'-end of the 1st intron of the *Cfa_SVIP* gene, respectively, we conclude that the Cfa_15&SV mRNA had been created by a rare alternative splicing event (Fig. 1C).

Tissue distribution of mouse 15G11 and SVIP mRNAs. By accessing the dataset in the Mouse GeneAtlas GNF1M, MAS5, we confirmed the expression profiles of Mmu_15G11 and Mmu_SVIP mRNAs. The former was confirmed by searching with Mm.279725, and the latter, by searching with Mm.386823 (Table I). The Mmu_15G11 mRNA was expressed at high levels specifically in adult mouse testes, whereas the Mmu_SVIP mRNA was widely expressed in various organs and at especially high levels in the trigeminal nerve, spinal cord, dorsal root ganglia, testes, and prostate.

Searches with protein query sequences against 'Motif Libraries'. In order to identify the functions of 15G11 proteins, we searched motif databases at GenomeNet using the amino acid sequences of 15G11 and SVIP proteins as queries (see Materials and methods) (13,14).

By searching with protein sequences against the PROSITE profile library, we found one bipartite nuclear localization signal profile (NLS_BP) within the domains coded by the 3rd exon in mouse, rat, dog, bovine, and human 15G11 proteins (Fig. 3A) but not in their SVIP counterparts (data not shown). In eukaryotic cells, selective transport of proteins into the nucleus is mediated by short amino acid sequences (NLS_BP), which consist of two small clusters of basic residues separated by a linker sequence (Fig. 3A) (15,16). By searching with the protein sequences against the BLOCKS library, we also found several blocks of sequences showing weak similarity to the registered 'block motifs' within 15G11 proteins (Fig. 3B). We found 'block motifs' related to (i) porin and apical membrane antigen 1 signatures within the regions coded mainly by the 1st and 2nd exons, (ii) CTF/NF-I family-like sequences within the regions coded mainly by the 2nd and 3rd exons,

SPANDIDOS PUBLICATIONS

(A) Prosite profile: Bipartite nuclear localization signal profile (PS50079; NLS_BP)

Name	Position	Consensus: BBXXXXXXXXBBBBB	Score
Mmu_15G11	34..48	MCLRVKDEEPAQVYP EGPRRHHPSDVSTRQ SVEKRINYMQNLQKEKRKLGKRFARPNPIPDTGIL WT	4.0
Hsa_15G11	34..48	MCLYCWDIEPSQVNP EGPRQHHPSEVTERQ LANKRIQNMQHLKKEKRRLNKRFSRPSPIPEPGLL WSS	4.0
Rno_15G11	34..48	MCLRGRDEEPAQVYA EGPRRHYPSDVSTRQ SMEKRINYMQNLKKEKRKLGKRFARPTPIPDPGIL WP	4.0
Cfa_15G11	34..48	MCLCCTEDDAIQVNP EGPRWQHPSEVTARQ SRDKRIENMKNIRKEKRKFNKRFARPAPLPEPGLL WT	4.0
Bta_15G11	34..48	MCLCCREDDSVQVNP ESSSWRHPAEVTERQ STAKRIEHMRNLWRQKRKFNKRFSRPAPVPEPGLL WT	4.0

*** : :. ** . *.. ::*::*: ** ***: *::: ::**::.***:** *:*:.*:* *.

(B) Blocks identified within 15G11 proteins

Name	Position	Block Sequence	Score	Block Name
Hsa_15G11	7..44	DIePSqvnp eGPRqHhpsevterq lankrIqnmqhlkk	1106	IPB003394D
Hsa_15G11	7..31	DiEpSqvnp EgprqHhpsevtErq l	1148	IPB0032980
Bta_15G11	14..38	nP EsSsWrhpaevterq stakrieh	1178	IPB0032980
Mmu_15G11	12..62	qVYp egprrhHPsdvstrq sVekrinymqNLQkeKRKlgKrfarpnpipdt	1130	IPB000647A
Bta_15G11	12..62	qVnp eSSswrHPaevterq stakriehmrNLwrqKRKfnKrfsrpapvpEp	1100	IPB000647A
Rno_15G11	20..53	rhYpsDVstrq SmekrInyMQnlkKEkrklgKrf	1113	IPB005015D
Cfa_15G11	32..55	RDKRIENMknIRKeKRkfNkRfaR	1123	IPB002652A
Hsa_15G11	33..59	nkRiqnmQhlKkEkrrLnKRfSRpSpI	1204	IPB003417A
Rno_15G11	33..59	ekRinymQnlKkEkrkLgKRfaRpTpI	1147	IPB003417A
Cfa_15G11	39..48	MKNIRKEKRK	1131	IPB011575B

Block name = Description: IPB003394D = Porin, opacity type; IPB0032980 = Apical membrane antigen 1 signature; IPB000647A = CTF/NF-I family; IPB005015D = Vibrio thermostable direct hemolysin; IPB011575B = RasGAP; IPB002652A = Importin α-like protein, β-binding region; IPB003417A = Core binding factor, β subunit.

Figure 3. Prediction of functional motifs within the 15G11 proteins. (A) Multiple sequence alignment of 15G11 proteins and the identified PROSITE motif, bipartite nuclear localization signal profile (PS50079; NLS_BP) are shown. (B) Blocks were identified as described in Materials and methods within the 15G11 proteins. The other procedures, abbreviations, and symbols are as described in the Fig. 1 legend and/or Table I.

and (iii) importin α-like protein (ß-binding region), core binding factor (ß-subunit), and RasGAP-like sequences within the region coded by the 3rd exon (Fig. 3B).

Discussion

The phylogenetic tree shown in Fig. 2 revealed that an ancestral gene for mammalian *15G11* genes evolved from an ancestral gene for mammalian *SVIP* genes. Notably, the ancestral gene for the mammalian *SVIP* genes continued its evolution at a relatively stable pace after the duplication, whereas the ancestral gene for the *15G11* genes evolved quite rapidly (Fig. 2). These results suggest that after the duplication, the former has kept its function, while the latter lost its original function. After rapid sequence divergence, the ancestral gene for the *15G11* genes seemed to have acquired new function(s) and diverged further into the mammalian *15G11* genes (Fig. 2).

Since structure and function relationships of SVIP have been studied using rat SVIP (4), we tried to predict the functions of 15G11 by taking these results into consideration. Rat, as well as mouse, SVIP contains two coiled-coil regions, the N- and the C-terminal regions (Fig. 1B) (4). These two regions were prepared individually from rat SVIP, and tested as to whether they can interact with VCP/p97 (4). The N-terminal region covering 39 amino acid residues has been shown to bind with VCP/p97, whereas the C-terminal region, covering 37 amino acid residues, has no such activity (4). A novel VCP/p97-interacting motif (VIM) has been identified within the N-terminal 39 amino acid region of mouse, rat, and human SVIP proteins (see Fig. 1B, underlined amino acid residues in the Mmu_SVIP line) (5). Moreover, Mmu_SVIP possesses a potential myristoylation site (Gly2), and a palmitoylation site (Cys4) in the N-terminal region, and the N-terminal 9 amino acid residues including these two sites have been reported to function as a membrane anchor (Fig. 1B) (4).

On the other hand, the homology between mouse 15G11 and SVIP is relatively low within the first half of these proteins (Fig. 1B). We were not able to identify a consensus VIM within 15G11 proteins (Fig. 1B). Moreover, Mmu_15G11 does not possess the myristoylation site (Gly2), and the homologies between the N-terminal 9 amino acid residues of these two proteins are also relatively low (Fig. 1B). These results suggest that mouse 15G11 protein does not bind to VCP/p97 and that it has no membrane anchorage function. Notably, we found that all of the 15G11 proteins carry NLS_BP-like motifs (Fig. 3A), whereas none of the mammalian SVIP proteins carry the same motifs. These results suggest that 15G11 proteins are nuclear proteins but mammalian SVIP proteins are not.

Moreover, multiple alignments of 15G11 proteins revealed that the regions encoded by the 3rd exon were highly conserved (Fig. 3A), suggesting that these regions are important for the functions of 15G11 proteins. Interestingly, these regions contain 'block motifs' related to importin α-like protein (ß-binding region) and core binding factor (ß-subunit) (Fig. 3B); the former is related to the nuclear protein and the latter to protein-DNA interaction. Although the scores of these block-like sequences are not high, their distributions suggest that the region encoded by the 3rd exon not only has properties related to nuclear proteins but also has properties to interact with DNA (Fig. 3A and B).

As mRNA coding for Mmu_15G11 protein corresponded to one of the most highly expressed testis-specific mRNAs, we speculate that a significant amount of Mmu_15G11 protein is required in adult mouse testis nuclei, and that Mmu_15G11 proteins play pivotal role(s) for the maintenance of nuclear DNA structures within the adult mouse testes.

Acknowledgements

We thank Dr Etsuro Yamanishi, President of Hirakata Ryoikuen, for his constant support and encouragement. We thank the National Center for Biotechnology Informatics, USA; the Genomics Institute of the Novartis Research Foundation, USA; the Genome Bioinformatics Group, Center for Biomolecular Science and Engineering at the University of California Santa Cruz, USA; DNA Data Bank of Japan; GenomeNet, Kyoto University Bioinformatics Center, Japan; the Computational Biology Research Center, AIST, Japan, and Ensembl, the Wellcome Trust Sanger Institute, UK for access to the network servers.

References

1. Su AI, Cooke MP, Ching KA, *et al*: Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci USA 99: 4465-4470, 2002.
2. The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team: Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420: 563-573, 2002.
3. The FANTOM Consortium and RIKEN Genome Exploration Research Group and Genome Science Group: The transcriptional landscape of the mammalian genome. Science 309: 1559-1563, 2005.
4. Nagahama M, Suzuki M, Hamada Y, *et al*: SVIP is a novel VCP/p97-interacting protein whose expression causes cell vacuolation. Mol Biol Cell 14: 262-273, 2003.
5. Ballar P, Shen Y, Yang H and Fang S: The role of a novel p97/valosin-containing protein-interacting motif of gp78 in endoplasmic reticulum-associated degradation. J Biol Chem 281: 35359-35368, 2006.
6. Ballar P, Zhong Y, Nagahama M, Tagaya M, Shen Y and Fang S: Identification of SVIP as an endogenous inhibitor of endoplasmic reticulum-associated degradation. J Biol Chem 282: 33908-33914, 2007.
7. Bono H, Yagi K, Kasukawa T, *et al*: Systematic expression profiling of the mouse transcriptome using RIKEN cDNA microarrays. Genome Res 13: 1318-1323, 2003.
8. Altschul SF, Madden TL, Schaffer AA, *et al*: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402, 1997.
9. Kent WJ: BLAT - the BLAST-like alignment tool. Genome Res 12: 656-664, 2002.
10. Thompson JD, Higgins DG and Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680, 1994.
11. Saitou N and Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425, 1987.
12. Tamura K, Dudley J, Nei M and Kumar S: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599, 2007.
13. Falquet L, Pagni M, Bucher P, *et al*: The PROSITE database, its status in 2002. Nucleic Acids Res 30: 235-238, 2002.
14. Henikoff JG, Greene EA, Pietrokovski S and Henikoff S: Increased coverage of protein families with the blocks database servers. Nucleic Acids Res 28: 228-230, 2000.
15. Dingwall C and Laskey RA: Nuclear import: a tale of two sites. Curr Biol 8: R922-R924, 1998.
16. Makkerh JP, Dingwall C and Laskey RA: Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. Curr Biol 6: 1025-1027, 1996.