Gene 517 (2013) 147–157

Contents lists available at SciVerse ScienceDirect

Gene

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

Structural and phylogenetic comparison of napsin genes: The duplication, loss of function and human-specific pseudogenization of napsin B

Kazuhisa Nishishita *, Eiko Sakai, Kuniaki Okamoto, Takayuki Tsukuba

Division of Oral Pathopharmacology, Department of Developmental Reconstructive Medicine, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki 852-8588, Japan

ARTICLE INFO

Article history:
Accepted 4 January 2013
Available online 17 January 2013

Keywords:
Napsin
Phylogeny
Synteny
Pseudogenization
SNP
Gene duplication

ABSTRACT

Aspartic proteinases form a widely distributed protein superfamily, including cathepsin D, cathepsin E, pepsins, renin, BACE and napsin. Human napsin genes are located on human chromosome 19q13, which comprises napsin A and napsin B. *Napsin B* has been annotated as a pseudogene because it lacks an in-frame stop codon; its nascent chains are cotranslationally degraded. Until recently, there have been no studies concerning the molecular evolution of the napsin protein family in the human genome. In the present study, we investigated the evolution and gene organization of the napsin protein family. *Napsin B* orthologs are primarily distributed in primates, while *napsin A* orthologs are the only napsin genes in other species. The corresponding regions of *napsin B* in the available sequences from primate species contain an in-frame stop codon at a position equivalent to that of human *napsin A*. In addition, a rare single-nucleotide polymorphism (SNP) that creates a proper stop codon in human *napsin B* was identified using HapMap populations. Recombinant protein expression and three-dimensional comparative modeling revealed that napsin B exhibits residual activity toward synthetic aspartic protease substrates compared with napsin A, presumably through a napsin B-specific Arg287 residue. Thus, *napsin B* was duplicated from *napsin A* during the early stages of primate evolution, and the subsequent loss of napsin B function during primate evolution reflected ongoing human-specific *napsin B* pseudogenization.

1. Introduction

Aspartic proteinases are acidic proteolytic enzymes that have a bilobal structure with two domains (Hsu et al., 1977). The active site of the aspartic proteinase contains two aspartate residues positioned in the middle of a cleft between the N- and C-terminal domains of the molecule and is partially covered by a hairpin loop, termed the flap or S1 subsite, protruding from the N-terminal domain (Sepulveda et al., 1975). Aspartic proteinases form a multigenic family that is widely distributed in organisms, and the major members of this family can be arranged into distinct clusters of orthologous groups, including cathepsin D, cathepsin E, nothepsin, renin, BACE, pepsin, and the fetal forms, pepsin Y and pepsin F (Borrelli et al., 2006; Carginale et al., 2004; Kageyama, 2002; Vassar et al., 1999; Xin et al., 2000).

The human napsin A (*NAPSA*) transcript is primarily expressed in the lung and kidney, but minor expression of *NAPSA* has also been observed in the prostate, connective tissue and the eye. *NAPSA* is expressed in alveolar type II cells and well-differentiated lung adenocarcinomas, whereas *NAPSA* expression is weak in poorly differentiated tumors, making this protein a promising diagnostic marker for primary lung adenocarcinomas (Chuman et al., 1999; Dejmek et al., 2007; Hirano et al., 2003; Ueno et al., 2008). It has also been shown that NAPSA protein is present in human urine. The napsin B (NAPSB) transcript is predominantly expressed in blood and lymphoid tissues, such as tonsil, lymph node, bone marrow, and spleen. In humans, the *NAPSA* and *NAPSB* genes are tandemly located on chromosome 19q13. The napsin genes contain nine exons, and *NAPSA* and *NAPSB* have the same exon organization. Rodents express a single napsin, designated napsin A (Napsa). Thus, Tatnell et al. (1998) proposed that *NAPSA* and *NAPSB* were derived from a relatively recent gene duplication event, in evolutionary terms, after the divergence of mice and humans, though the exact duplication timing and process are unknown until now. *NAPSB* has been annotated as a pseudogene because it lacks an in-frame stop codon at a position equivalent to that of human *NAPSA* (Tatnell et al., 1998). Recently, it has been reported that chimpanzee *NAPSB* contains an in-frame stop codon, suggesting that chimpanzee *NAPSB* encodes a functional aspartic

Abbreviations: NAPSA, napsin A; NAPSB, napsin B; SNP, single-nucleotide polymorphism; cDNA, DNA complementary to RNA; bp, base pair(s); PCR, Polymerase Chain Reaction; kb, kilobase(s) or 1000 bp; NJ, Neighbor-Joining; ML, Maximum-Likelihood; KCNC3, potassium voltage gated channel Shaw-related subfamily member 3; NR1H2, nuclear receptor subfamily 1 group H member 2; atf4, activating transcription factor 4; smcr7l, Smith–Megenis syndrome region candidate 7-like; G418, Geneticin; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide-gel electrophoresis; EDTA, ethylenediaminetetraacetic acid; kDa, kilodalton(s); KLH, Keyhole limpet hemocyanin.

* Corresponding author. Tel.: +81 95 819 7654; fax: +81 94 819 7655.
E-mail addresses: kazu@nagasaki-u.ac.jp (K. Nishishita), eiko-s@nagasaki-u.ac.jp (E. Sakai), k-oka@nagasaki-u.ac.jp (K. Okamoto), tsuta@nagasaki-u.ac.jp (T. Tsukuba).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.01.013

protease (Puente et al., 2005). The reference allele at polymorphic sites has been defined using the most common allele obtained from the alignment of multiple individual genome sequences. However, the major allele encodes a loss-of-function variant, reflecting an inherent bias toward annotating functional genes in the reference genome; thus, events that might lead to gene inactivation are largely overlooked in automatic annotation processes (Balasubramanian et al., 2011). In this context, *NAPSB* orthologs in other species are typically annotated as *NAPSA*-like genes, and no *NAPSB* orthologs have been described in other species. Here, we investigated the evolution and gene organization of the two napsins, *NAPSA* and *NAPSB*, and traced the evolutionary origin of the subfamily of these genes in animals. Moreover, we compared the enzymatic activity and three-dimensional structure of the NAPSA and NAPSB proteins.

2. Materials and methods

2.1. In silico analyses

2.1.1. Identification of NAPSB genes

Napsins and related aspartic protease sequences were identified in the Ensembl and GenBank databases for the following species with available genome sequences: *Homo sapiens* (human), *Pan troglodytes* (common chimpanzee), *Pan paniscus* (bonobo), *Pongo abelii* (orangutan), *Macaca mulatta* (rhesus monkey), *Nomascus leucogenys* (gibbon), *Otolemur garnettii* (galago), *Sus scrofa* (pig), *Bos taurus* (cow), *Equus caballus* (horse), *Ailuropoda melanoleuca* (giant panda), *Canis lupus familiaris* (dog), *Mus musculus* (mouse), *Rattus norvegicus* (rat), *Ornithorhynchus anatinus* (platypus), *Gallus gallus* (chicken), *Xenopus*

Table 1
List of accession numbers for all aspartic protease sequences used in the phylogenetic analysis.

Organism	Name	Accession number	
		Protein	mRNA
Homo sapiens	HsaNAPSA	NP_004842.1	NM_004851.1
	HsaNAPSBp		NR_002798.1
	HsaCTSE	NP_001901.1	
	HsaCTSD	NP_001900.1	
	HsaREN	NP_000528.1	
	HsaPEPC	NP_002621.1	
	HsaPEPA4	NP_001073276.1	
Pan troglodytes	PatNAPSA	XP_524345.2	XM_524345.3
	PatNAPSB	XP_530061.2	XM_530061.3
Pan paniscus	PapNAPSB		XM_003813624.1
Pongo abelii	PoaNAPSB		XM_002829607.2.2[a]
Nomascus leucogenys	NolNAPSA		XM_003269801.1
	NolNAPSB	XP_003269848.1	XM_003269800.1
Macaca mulatta	MamNAPSA	XP_001116026.1	ENSMMUT00000018797
	MamNAPSB	ENSMMUP00000031507	ENSMMUT00000038406
Callithrix jacchus	CajNAPSA		XM_003735639.1
	CajNAPSB		XR_144592.1
Otolemur garnettii	OtgNAPSA	XP_003801583.1	
Sus scrofa	SusNapsA		XM_003127363.2
Bos taurus	BotNapsA		XM_002695127.1
Equus caballus	EqcNapsA		XM_001490835.1
Ailuropoda meranoleuca	AimNapsA	XP_002917936.1	
Canis lupus famioliaris	CafNapsA	XP_533610.2	XM_533610.3
Mus musculus	MumNapsA	NP_032463.1	NM_008437.1
	MumPepA5	NP_067428.2	
	MumRen1	NP_112469.1	
Rattus norvegicus	RnoNapsA		NM_031670.2
Ornithorhynchus anatinus	OanNapsA	ENSOANP00000019807	
Gallus gallus	GagPep	ENSGALP00000000593	
	GagCtsE	ENSGALP00000001138	
	GagCtsD	ENSGALP00000010662	
Xenopus laevis	XlaNapsA	NP_001083566.1	
	XlaCtsE	BAC57453.1	
Xenopus tropicalis	XtrNapsA	NP_001005701.1	
Latimeria chalumnae	LacNapsA	ENSLACP00000016743	
Takifugu rubripes	TruPep	NP_001072051.1	
	TruCtsD1	NP_001072052.1	
	TruCtsD2	NP_001072053.1	
	TruRen	NP_001072054.1	
	TruNts	NP_001072055.1	
Chionodraco hamatus	ChhNts	CAA11580.1	
	ChhCtsD	CAA07719.1	
Clupea harengus	ClhCtsD	AAG27733.1	
Danio rerio	DarNapsA	AAH56836.1	
	DarCtsD	NP_571785.1	
	DarRen	AAO31713.1	
	DarNts	NP_571879	
Oryzias latipes	OrlNapsA	ENSORLP00000016894	
Sparus aurata	SpaCtsD	AAB88862	
Haemonchus contortus	HacPep	CAA96571.1	

[a] This sequence is annotated as *NAPSA* in the GenBank.