Available online at www.sciencedirect.com
SCIENCE DIRECT
Genomics 87 (2006) 139 – 150
GENOMICS
www.elsevier.com/locate/ygeno

Bioinformatic characterization of the SynCAM family of immunoglobulin-like domain-containing adhesion molecules

Thomas Biederer *

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

Received 25 May 2005; accepted 15 August 2005
Available online 28 November 2005

Abstract

SynCAM 1 (synaptic cell adhesion molecule 1, alternatively named Tslc1 and nectin-like protein 3) belongs to the immunoglobulin superfamily and is an adhesion molecule that operates in a variety of important contexts. Exemplary are its roles in adhesion at synapses in the central nervous system and as tumor suppressor. Here, I describe a family of genes homologous to SynCAM 1 comprising four genes found solely in vertebrates. All SynCAM genes encode proteins with three immunoglobulin-like domains of the V-set, C1-set, and I-set subclasses. Comparison of genomic with cDNA sequences provides their exon–intron structure. Alternative splicing generates isoforms of SynCAM proteins, and diverse SynCAM 1 and 2 isoforms are created in an extracellular region rich in predicted O-glycosylation sites. Protein interaction motifs in the cytosolic sequence are highly conserved among all four SynCAM proteins, indicating their critical functional role. These findings aim to facilitate the understanding of SynCAM genes and provide the framework to examine the physiological functions of this family of vertebrate-specific adhesion molecules.

Keywords: Cell adhesion molecules; Immunoglobulin superfamily; SynCAM; Synaptic cell adhesion molecule; IGSF4; Tslc; Nectin-like molecule; SgIGSF; Central nervous system; Surface recognition; Synapse formation; Genomic structure; Exon–intron junctions

Members of the immunoglobulin (Ig) superfamily are a large group of proteins with diverse roles in extracellular recognition [1]. SynCAM 1 (synaptic cell adhesion molecule 1; human gene symbol *IGSF4*) belongs to this superfamily and is a membrane protein that functions as intercellular adhesion molecule in several tissues, including the brain. SynCAM 1 contains three extracellular Ig-like domains, a single transmembrane sequence, and a short cytosolic tail sporting two identified protein–protein interaction motifs.

Different SynCAM family members have been identified by a number of independent approaches. SynCAM 1 serves as synaptic adhesion molecule in the central nervous system and is expressed throughout the brain [2]. It is localized to synaptic plasma membranes, where it engages in an interaction that bridges the synaptic cleft [2]. Notably, SynCAM 1 is sufficient to induce neurons to form functional presynaptic specializations and promotes synaptic transmission [2,3]. All activities of SynCAM 1 identified in neurons depend on its extracellular, Ig-like domain-mediated interactions [2,3]. SynCAM 1 was also reported to mediate the attachment of mast cells to neurites [4]. Independently, SynCAM 1 was mapped and cloned as a potential tumor suppressor gene [5,6]. This role was confirmed by analysis of SynCAM 1-expressing tumor cells in immunocompromised mice, leading to the alternative naming of SynCAM 1 as tumor suppressor in lung cancer 1 (Tslc1) [7]. In addition, SynCAM 1 was cloned from PC12 cells as a retinoic-acid-inducible gene, implicating it in the differentiation of these cells to a neuron-like morphology [8], and as an Ig-domain-containing protein expressed in testis during spermatogenesis, leading to its naming as SgIGSF (spermatogenic immunoglobulin superfamily member) [9]. SynCAM 1 transcripts are abundant in brain, present in testis, and also more weakly detected in other tissues ([9,10] and T. Biederer et al., unpublished observations). A different family member, SynCAM 3, is expressed in epithelia and was named nectin-like protein 1 (necl-1) due to its overall domain organization, which is similar to that of nectin [11]. Functional roles for SynCAMs 2 and 4 remain to be identified.

* Fax: +1 203 785 6404.
E-mail address: thomas.biederer@yale.edu.

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.08.017

A protein family relevant to the SynCAM adhesion molecules comprises the above-mentioned nectins, a group of four immunoglobulin superfamily members (reviewed in [12,13]). Nectins are expressed in a variety of tissues and are involved in intercellular adhesion in epithelial and neural tissues, most prominently at adherens junctions. SynCAM and nectin proteins exhibit an overall similar domain organization with three extracellular Ig-like domains, a single transmembrane region, and a carboxy-terminal intracellular tail. However, SynCAM and nectin proteins are not homologous. They share only a low sequence identity, and their Ig-like domains are not related (data not shown). Furthermore, the intracellular tails of SynCAM proteins, characteristic of this protein family (see below), are not similar to those of nectins. Interestingly, the extracellular domains of SynCAM 1 and nectin 3 interact with each other, as well as the extracellular domains of SynCAM 3 and nectins 1 and 3 [11,14]. Functional roles of these interactions remain to be characterized.

In light of the important functions SynCAM proteins exert in different physiological contexts ranging from the differentiation of synapses in the brain to tumor suppression, it is necessary to characterize this gene family. I here show that the SynCAM family consists of four conserved members. All SynCAM proteins display the same domain organization as SynCAM 1. Genes encoding SynCAM family members are found throughout vertebrate genomes from puffer fish to humans, but are absent in invertebrates. While SynCAM protein sequences are relatively conserved, the genes encoding them display remarkable size differences up to 60-fold in humans and mice. Furthermore, three of the four genes encoding SynCAM family members undergo alternative splicing in humans and mice, but the number and location of alternatively spliced exons vary between them. Particularly, the analysis of splice patterns draws attention to the juxtamembranous region of the extracellular domain of SynCAM 1 and 2 and potential functional roles of this sequence and its splicing. In addition, this study shows that the carboxy-terminal intracellular tails of SynCAMs are the most conserved sequences, indicating the critical role of cytosolic protein–protein interaction motifs for the functions of SynCAMs. This analysis aims to further our understanding of the roles of SynCAM proteins in the differentiation of tissues such as the developing brain and as tumor suppressors.

Results

The SynCAM family comprises four genes, encoding proteins with three Ig-like domains, a single transmembrane region, and a short cytosolic tail with a protein 4.1 interaction sequence and a PDZ type II motif (see below for an analysis of these genes and their sequence features). Different names are used for individual SynCAM family members (Table 1 and references therein). To clarify their nomenclature, I propose to name these proteins SynCAM 1, 2, 3, and 4 for the following reasons. First, the high degree of conservation on the amino acid level warrants the classification of SynCAM proteins as a family (see below). Second, this classification as gene family is independently supported by the highly similar exon usage of the four SynCAM genes. Third, this nomenclature was used in the first published reference to this protein family [2]. Fourth, the nomenclature as SynCAMs proposed here is the only homology-based classification that includes all family members. Fifth, while it has been proposed to refer to these proteins as nectin-like molecules or nectin-like proteins (necl) based on similarities in their domain organization [15], this does not appear fitting on homology grounds, as that classification includes two nonhomologous proteins (the poliovirus receptors necl-5 and necl-6 [15]). Last, the gene symbol *IGSF4* (immunoglobulin superfamily member 4) appears neither useful nor fitting for naming the protein. It is not indicative of a function, and IGSF 1, 2, and 3 are not homologous to the SynCAM family. I therefore propose that these four genes are referred to as the SynCAM family, based on their homology to SynCAM 1.

Table 1
Genes and nomenclatures

Human gene symbol	*IGSF4*	*IGSF4D*	*IGSF4B*	*IGSF4C*
Mouse gene symbol	*Igsf4a*	*Igsf4d*	*Igsf4b*	*Igsf4c*
Protein name	SynCAM 1 [2]	SynCAM 2 [2]	SynCAM 3 [2]	SynCAM 4 [2]
Alternate protein names	Necl-2 [14] Tslc1 [7] SgIGSF [9] RA175 [8]	Necl-3 [15]	Necl-1 [11] Tsll1 [45]	Tsll2 [45]

Human and mouse genes beginning with *IGSF4* and *Igsf4*, respectively, encode SynCAM proteins. Different names used for individual SynCAM proteins are listed. Ensembl gene IDs for *IGSF4*, *IGSF4B*, *IGSF4C*, and *IGSF4D* are ENSG00000182985, ENSG00000175161, ENSG00000162706, and ENSG00000105767, respectively. The Ensembl gene IDs for *Igsf4a*, *Igsf4b*, *Igsf4c*, and *Igsf4d* are ENSMUSG00000032076, ENSMUSG00000064115, ENSMUSG00000005338, and ENSMUSG00000054793.

The SynCAM family of membrane proteins comprises four members

SynCAM sequences and splice products were identified in database searches as described under Methods and are listed in Supplementary Table S1. Four human SynCAM family members can be identified that share a high degree of amino acid identity, ranging for the full-length proteins from 50% (SynCAM 2 vs 3) to 36% (SynCAM 1 vs 3 or 4). The significant conservation of their three extracellular Ig-like domains is apparent in the alignment of their sequences (Fig. 1; see below for an analysis). In addition, this alignment highlights the striking similarity of the intracellular carboxy-terminal sequences of SynCAM proteins (Fig. 1) and their conserved intracellular interaction motifs with protein 4.1 and PDZ type II domains (see below).

A phylogenetic comparison of human and mouse SynCAM full-length sequences classifies SynCAMs 2 and 3 as more closely related to each other than to SynCAM 1 (Fig. 2). SynCAM 4 is a more distant family member. This interpretation is confirmed by a detailed analysis of their individual Ig-like domains (see below). All human SynCAM family