Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 359 (2005) 63 – 72
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Identification, expression analysis, genomic organization and cellular location of a novel protein with a RhoGEF domain

Sze-Wah Tse [a], Jennifer A. Broderick [a], Mei-Ling Wei [a], Min-Hua Luo [a,b], Deborah Smith [a], Peter McCaffery [a,c], Stefan Stamm [d], Athena Andreadis [a,c,*]

[a] Department of Biomedical Sciences, E.K. Shriver Center for Mental Retardation, Waltham, MA 02452, United States
[b] Department of Microbiology, Xiangya Medical School, Central South University, Changsha 410078, Hunan, P.R. China
[c] Department of Cell Biology, University of Massachusetts Medical School, Worcester, MA 02116, United States
[d] University of Erlangen-Nurenberg, Institute of Biochemistry, 91054 Erlangen, Germany

Received 21 March 2005; received in revised form 17 May 2005; accepted 3 June 2005
Received by A.J. van Wijnen

Abstract

In this study we describe the identification and characterization of a novel cytosolic protein of the guanine exchange factor (GEF) family. The human cDNA corresponds to predicted human protein FLJ00128/FLJ10357 located on chromosome 14q11.2. The deduced protein sequence contains in its C-terminus a RhoGEF domain followed by a pleckstrin domain. Its N-terminus, central region and RhoGEF/pleckstrin domain are homologous to the recently identified zebrafish Quattro protein, which is involved in morphogenetic movements mediated by the actin cytoskeleton. Based on the homology of our protein's RhoGEF domain to the RhoGEF domains of Trio, Duo and Duet and its homology with Quattro, we named it Solo. The Solo mRNA is ubiquitously expressed but enriched in brain, its expression peaks perinatally and it undergoes extensive alternative splicing. In both myoblasts and neuroblastoma cells, the Solo protein is concentrated around the nucleus.

Keywords: Yeast two-hybrid screen; FLJ protein; Alternative splicing; Perinuclear localization

1. Introduction

Guanine exchange factors (GEF) stimulate Rho and Rac signal transduction molecules by switching them from the inactive (GDP-bound) to the active (GTP-bound) form. These molecules are often involved in organizing the cytoskeleton and act as axon guidance molecules (Schmidt and Hall, 2002). In a yeast two-hybrid screen aimed to identify proteins interacting with the hinge domain of tau, we isolated a cDNA corresponding to predicted human protein FLJ00128/FLJ10357. Here we report the genomic organization, expression analysis and cellular localization of this novel gene and its product, which we named Solo to reflect its homology to Duo, Duet, Trio and Quattro (members of the RhoGEF superfamily; Erickson and Cerione, 2004). The C-terminus of the deduced Solo protein sequence contains a RhoGEF domain followed by a pleckstrin domain, and three of its regions are homologous to equivalent regions of the zebrafish Quattro protein, which is involved in morphogenetic movements mediated by the actin cytoskeleton (Daggett et al., 2004). The Solo transcript varies in different tissues and developmental stages and also undergoes extensive alternative splicing, which influences the length of its reading frame.

2. Materials and methods

2.1. Yeast two-hybrid assay

We used the Clontech GAL4 yeast two-hybrid system to identify proteins interacting with the hinge region of tau.

Abbreviations: EST, expressed sequence tag; GEF, guanine exchange factor; PH, pleckstrin domain.

* Corresponding author. Shriver Center, 200 Trapelo Road, Waltham, MA 02452, United States. Tel.: +1 781 642 0279; fax: +1 781 642 0017.
E-mail address: Athena.Andreadis@umassmed.edu (A. Andreadis).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.06.025

The bait contained human tau exons 5 and 6 (56 and 198 nt, respectively) and the first 80 nucleotides of tau exon 7. To prepare it, we inserted a 230 bp EcoRI/NotI tau cDNA fragment into vector pGBT9 (Clontech). Afterwards, we filled in the EcoRI site to create an AseI site, so that the fusion protein would have the correct reading frame in its tau portion.

The 5/6/7 bait and a rat brain (day P6) library in vector pGAD424 (Clontech) were serially transformed into yeast reporter strain HF7c, which was plated on SD plates lacking tryptophan, leucine and histidine (the reporter gene). The clones which grew in the absence of histidine were further tested for lacZ activity (development of blue color on plates containing X-gal). The library plasmids contained in the positive yeast colonies were isolated by passage through the leuB$^-$ *E. coli* strain HB101 (selected for leucine prototrophy on M9 minimal medium). Plasmid DNA was produced by Qiagen Tip-50 s and sequenced in an ABI fluorescent sequencer using either primers specific to GAD424 or primers from the Solo sequence.

2.2. cDNA cloning, sequencing and sequence analysis

We isolated two unique clones from the two-hybrid screen. One codes for TIF1, which is involved in transcriptional regulation (Le Douarin et al., 1997). The other, which we originally named mp31, codes for a novel protein with RhoGEF and pleckstrin domains at its C-terminal region. For reasons described in Results, we eventually decided to call this molecule Solo.

Since yeast hybrid systems contain fusion proteins which de facto lack the 5′end, we used the 5′most 300 nucleotides of mp31 to screen a rat testis library (Clontech). This yielded three clones, 16-1, 6-7 and 1-17; the latter contained a significant length of additional 5′sequence past mp31. We then used the 5′end of 1-17 to search two cDNA libraries, rat testis (Clontech) and rat adult brain (Stratagene) for clones extending 5′wards. The screenings yielded a cDNA, Solo 1-17, which minimally extended Solo. At that point, no rat libraries yielded any more clones extending upstream; however, a human EST from the Japanese HUGO base clearly showed a longer ORF. So we used PCR techniques to probe a mouse cerebellum cDNA library (Incyte) which gave us a short additional 5′end sequence.

At that point, the genomic version of Solo became available for both human and mouse. We used the combined cDNA sequence of all the clones to search for the corresponding mouse and human genomic sequence in the GenBank database. Alignment analysis, based on loss of contiguity between the cDNA and genomic segments and on the presence of typical 5′ and 3′ splice site motifs, was carried out to determine the exon/intron boundaries (shown in Table 1). Comparisons of the two genomic

Table 1
Solo exon/intron organization—mouse (human)

Exon	C	3′ss cagG	Exon length (bp)	5′ss CAGgtragt	C	Intron length (kbp)	Comments/features
1	?	?	>110	ATGgtgagt	7/9	2.6	ATG
2	4/4	cagG	198	TGTgtgagt	6/9	1.2 (0.7)	
3*	4/4	cagG	1237 (1249)*	CAGgtgaga	8/9	0.11*	
4	4/4	cagG	168	CTGgtcagt (CGGgtcagt)	7/9	0.75 (0.15*)	
5	4/4	cagG	127 (121)	TAGgtaagc (CAGgtaacc)	7/9	0.6	
6*	4/4	cagG	97	CAGgtaatc (CAGgtaacc)	7/9	0.11*	
7*	4/4 (3/4)	cagG(tagG)	81	CAGgttggt (CAGgttgag)	7/9 (5/9*)	0.12*	
8	4/4	cagG	117	CAGgtaaac (CAGgtaagc)	7/9 (8/9)	2.0 (1.3)	
9*	4/4	cagG	96	GAGgtatgg (GAGgtatga)	6/9	0.09* (0.11*)	
10	4/4	cagG	116	CAAgtaaga (CAAgtacga)	7/9 (6/9)	0.5 (0.4)	
11	4/4 (3/4)	cagG (aagG)	127	CAGgtaaag (CAGgtgagc)	7/9 (8/9)	1.1 (1.65)	
12*	4/4	cagG	123	CAGgtaagg (CAGgtgaga)	8/9	0.08* (0.9)	
13	4/4	cagG	144	CAGgtgagg (CAGgtgcat)	8/9 (6/9)	0.5	
14*	3/4	tagG	611	CAGgtgaga	8/9	0.12*	
15	3/4	cagC (cagT)	235	CATgtgagc (CACgtgagt)	7/9 (8/9)	0.3 (0.35)	
16	4/4	cagG	87	AAGgtaaac (AAGgtaact)	6/9 (7.9)	>2.3 (0.9)	
17	3/4	tagG	216	GAGgtaagg (GAGgtgagg)	7/9	0.8 (0.7)	
18	3/4	cagA	171	AAGgtgtgc (AAGgtacga)	6/9	1.3 (0.8)	
19	3/4	cagA	178	AGGgtaacg (AGGgtactg)	5/9	1.2 (1.15)	
20	2/4* (3/4)	tagA (cagA)	105	GAGgtgagg	7/9	0.2	
21	3/4	cagC	144	TGGgtgagg	6/9	0.2	
22a*	2/4*	tagC	132 (135)	CTTgttagg (TTGgtaggg)	5/9*	0.15*	TAA (CAA)
22b	3/4	cagC	131	CAGgtgagt (CAGgtaagt)	9/9	0.25 (0.7)	
Mini*	3/4	tagG	21*	AAGgtcaga	5/9*		Functional 3′ss only in mouse
23	3/4	cagA	47	TAGgtgagt (TGGgtgagt)	8/9 (7/9)	0.3	TGA
24	4/4	cagG	>300	?			

The items in parentheses show the properties of the human exon, if different from the mouse.
C=agreement of splice sites with the consensus sequence.
*=unusual features (very short or long exon, very short intron, very low splice site agreement with consensus).