Genomic organization, transcription start sites, and chromosomal location of the Drosophila cortactin gene.

Takanori Katsube†*, Naoko Hashimoto, Yuka Sasaki, Atsushi Yamashita, Manabu Takahisa, Ryu Ueda, Shin Togashi††

Neurogenetics Research Project, Mitsubishi Kasei Institute of Life Sciences, 11, Minami-ooya, Machida-shi, Tokyo 194-8511, Japan

(Received 3 July 2000, accepted 26 January 2001)

An actin filament binding protein cortactin was initially identified as a major phosphotyrosine-containing protein in v-Src-transformed chicken embryo fibroblast cells. The mouse, human, and Drosophila homologs were independently identified as a signaling molecule involved in a mitogenic response, as a product of a putative oncogene EMS1, and as a molecule interacting with a scaffolding protein ZO-1, respectively. In this report, we describe the cloning of the Drosophila cortactin gene, which consists of four exons and three introns, covering 3 kilobases in length. All exon-intron junctions are well matched with the GT/AG consensus sequence. S1 nuclease mapping revealed one major and several minor transcription start sites. The cytological location of the Drosophila cortactin gene is between chromosome segments 93B3 and 93B7.

INTRODUCTION

Cortactin, originally identified as a prominent substrate for Src protein tyrosine kinase, is a filamentous actin binding protein (Wu & Parsons, 1993; Wu et al., 1991). It also has a filamentous actin cross-linking activity, which is down-regulated upon phosphorylation mediated by Src (Huang et al., 1997). The human homolog is encoded by a putative oncogene, EMS1. Its participation in the invasion and metastasis of cancer cells was supported by studies on cultured cells, which are transfected with cortactin expression vectors (Huang et al., 1998; Patel et al., 1998; Schuuring et al., 1993). Thus, cortactin is implicated in some signaling pathways that regulate cytoskeleton dynamics. However, the molecular bases for these processes remain unclear.

Recently, we identified a Drosophila homolog of cortactin (DCortactin) as a molecule associating with Drosophila ZO-1, a component of cell-cell junctions (Katsube et al., 1998; Takahisa et al., 1996). These proteins predominantly colocalize at cell-cell junction sites in epithelial cells of Drosophila imaginal discs. We also demonstrated an association between cortactin and ZO-1 in mouse. In addition, immunohistochemical analysis of human and mouse tissues has revealed that mammalian cortactin is concentrated at the most apical region of the polarized epithelium, where tight junctions are formed (Wu & Montone, 1998). The association with ZO-1 may be significant for cortactin to participate in its cellular function. Determination of the primary structure as well as the cytological location of the DCortactin gene lays a foundation for further analysis using molecular genetics in Drosophila, which should provide new insights into the cellular functions of cortactin. We now report the molecular cloning of the DCortactin gene.

MATERIALS AND METHODS

1. Isolation of genomic clones and nucleotide sequencing

DCortactin cDNA probes were prepared from a plasmid clone pCT407 (Katsube et al., 1998). The 2.4-kilobase (kb) cDNA insert was divided into 3 fragments, which are the 5'-end 0.9-kb, central 0.9-kb, and 3'-end 0.6-kb probes. The probes were labeled with $^{32}$P by random priming. $3 \times 10^5$ clones of a Canton-S genomic library in EMBL3 SP6/T7 (Clontech Laboratories, Palo Alto, CA, USA) were screened with the 5'- and 3'-end cDNA probes. Restriction fragments of some positive phage clones were subcloned into pBluescript II SK+ (Stratagene, La Jolla, CA, USA) and sequenced using an ABI Prism 377 DNA Sequencer (Perkin-Elmer Co., Foster...
2. S1 nuclease protection assay A single stranded DNA probe was prepared by an asymmetric polymerase chain reaction (PCR) using the plasmid pgT407B4.0, which is a subclone of a 4.0-kb BamHI genomic DNA fragment from the DCortactin locus, as a template. 10 ng of pgT407B4.0 was digested with BamHI and added to 50 µl of PCR reaction buffer containing 100 µM each of dNTP and 7 pmol of DcoPex2 primer 5'-end labeled with 32P. The reaction was processed for 30 cycles at 94°C for 30 s, 50°C for 30 s, and 72°C for 2 min. Total RNA samples were isolated from 0- to 12-h embryos, 3rd instar larvae, adult male, and adult female flies using RNeasy (Qiagen, Hilden, Germany). 20 µg of total RNA was hybridized to 30 µl of the single stranded DNA probe and treated with S1 nuclease essentially as described (Sambrook et al., 1989). In some experiments, 200 µg of total RNA was further purified with an Oligotex-dT30 column (Takara-shuzou, Kyoto, Japan) and resulting poly(A)+ RNA was processed for S1 nuclease protection analysis as above. The protected fragments were separated on a sequencing gel (7% Long Ranger, FMC Bioproducts, Rockland, ME, USA) and analyzed by autoradiography.

3. Rapid amplification of cDNA 5' ends (5' RACE) Poly(A)+ RNA samples from 0- to 12-h embryos and 3rd instar larvae were reverse-transcribed by SuperScript II reverse transcriptase (Gibco BRL, Rockville, MD, USA) using the oligonucleotide Dcort432 or random hexamers as primers. Second strand synthesis and adaptor ligation were carried out using a Marathon cDNA Amplification kit (Clontech Laboratories). The resulting cDNA products were subjected to a polymerase chain reaction using an adaptor primer AP1 (Clontech Laboratories) and the oligonucleotide primer Dcort415. Since a 49-nucleotide adaptor sequence is added to the 5' end of each amplified product, the calculated size of the product for the previously reported DCortactin cDNA is 420 base pairs. The sequences of the oligonucleotide primers were as follows: Dcort432, 5'-ttcttgtccagatcggcctg-3'; Dcort415, 5'-ctgctcttcgtgacaggtcgcctgg-3'.

4. Reverse transcription and polymerase chain reaction (RT-PCR) RT-PCR was conducted as previously described (Katsube et al., 1997). The nucleotide sequences of the primers were 5'-caggccgatctggacaagaa-3' and 5'-caggccgatctggacaagaa-3'.

5. In situ hybridization to polytene chromosomes Salivary gland polytene chromosomes were prepared from the wandering third instar larvae of a wild type strain, Canton-S. The probe was labeled with biotin-11dUTP by nick translation. The hybridized probes were detected using an in Situ Hybridization and Detection System (Gibco BRL).

6. Southern blot analysis Genome DNA was isolated from Canton-S wild type or deficiency line (Df(3R)e-RI, Df(3R)e-BS2, Df(3R)e-N19, or Df(3R)e-F1) adult flies as previously described (Pirrotta, 1986). EcoRI digested genome DNA samples were separated on a 0.7% agarose gel and transferred to a Hybond-N+ nylon membrane (Amersham Life Science, West Chester, PA, USA). The DCortactin cDNA central probe described in section 1 in MATERIALS AND METHODS was hybridized to the membrane and detected by autoradiography. To measure the intensity of hybridized bands, a digitized image of the autoradiograph was analyzed using an Image Gauge ver. 3.4 software (Fuji Film, Tokyo, Japan).

RESULTS AND DISCUSSION

1. Isolation and organization of the DCortactin gene Southern blot analyses of Drosophila genome DNA digested with EcoRI were conducted using DCortactin

---

**Fig. 1.** Organization of DCortactin locus. Exon distribution of the DCortactin transcript is shown above the restriction map of the DCortactin locus. Boxes represent exons, and coding regions are filled. Cortactin repeats and the Src homology 3 domain coding regions are highlighted by hatched and dotted boxes, respectively. The thick bar represents a single stranded probe for an S1 nuclease protection assay. The arrow head represents an oligonucleotide primer for a 5' RACE. B, BamHI; E, EcoRI. Lines below the restriction map indicate representative phage clones obtained.

2. Transcription start sites of the DCortactin gene

S1 nuclease protection analysis using total RNA from Canton-S embryos revealed several transcription start sites of the DCortactin gene (Fig. 2). The major transcription start site is depicted by arrows in Fig. 2 and denoted by \(+1\) in Fig. 3. The minor transcription start sites are at nucleotide positions –30, –25, –13, and –4 with respect to the major one (indicated by asterisks in Figs. 2 and 3). S1 nuclease protection analyses using RNA samples from 3rd instar larvae, adult male, and adult female flies gave almost the same results (data not shown). It should be noted that the major transcription start site demonstrated by the S1 analysis is 7 base pairs downstream from the 5' end of the previously reported DCortactin cDNA (indicated by a closed rectangle in Fig. 3). In agreement with the result of the S1 analysis, a 5' RACE using an oligonucleotide primer complementary to the exon 2 upstream end exclusively amplified cDNA fragments of about 0.4 kb (data not shown, see section 3 in MATERIALS AND METHODS). For nucleotide sequencing, amplified fragments were cloned into a plasmid vector. All 36 clones sequenced have the 5' ends downstream from the nucleotide position –30, which is the most upstream transcription start site demonstrated above. Though, one clone is exceptional as it retain intron 1 but 2, and 5 clones have a 3-nucleotide insertion at the exon 1-2 junction site in the 5'-untranslated region. For the later cases, alternative splicing of intron 1 using a second acceptor site 3 base pairs upstream from the major one is most likely the cause. The AG consensus sequence is found beside the second acceptor site.

3. The DCortactin transcript is not likely to have isoforms

In the original reports on vertebrate cortactin, it was found to have 6.5 tandem 37-amino acid repeats, called cortactin repeats (Miglarese et al., 1994; Schuuring et al., 1993; Wu et al., 1991; Zhan et al., 1993). These cortactin repeats are necessary for binding to filamentous actin (Wu & Parsons, 1993). Transcripts encoding isoforms with 5.5 or 4.5 cortactin repeats have been also identified in mouse and rat (Ohoka & Takai, 1998; Wu &
Fig. 3. The nucleotide sequence around the transcription start sites of the DCortactin gene. The major transcription start site is referred to as nucleotide position +1. The minor transcription start sites demonstrated by the S1 nuclease protection analysis are indicated by asterisks and the 5' end of the previously reported cDNA by a filled rectangle. Introns 1 and 2, which are spliced out in most transcripts as well as in the previously reported cDNA, are represented by lower case letters. The second 3'-splice acceptor site of intron 1 is indicated by an open triangle. Positions of the oligonucleotide primers for the 5' RACE analysis (nucleotide position 590–618) and for the single stranded probe preparation of the S1 nuclease protection analysis (49–70) are underlined. Putative binding sites for transcription factors are double-underlined, although their physiological significance is not revealed.

Fig. 4. The DCortactin transcript has no isoforms differing in the cortactin repeats coding region. Poly(A)+ RNA samples from embryos (lane 2), larvae (lane 3), pupae (lane 4), and adult flies (lane 5) were subjected to RT-PCR using a primer set, which covers the cortactin repeats coding region. The products were separated on a 1.4% agarose gel. A 538-bp fragment corresponding to the transcript with 4 cortactin repeats was amplified from all samples tested. HindIII-digested φX174 phage DNA was loaded on lane 1 for a size reference.

4. Cytological location of the DCortactin gene

The DCortactin gene was mapped cytologically to the 93B region on the right arm of the third chromosome (Fig. 5 A). To confirm the chromosomal position of the gene, genomic Southern blot analysis of flies heterozygous for some deficiency chromosomes was performed using a DCortactin cDNA probe (Fig. 5 B). A 4.2-kb band is about half the intensity in Df(3R)e-R1 heterozygotes compared to that in other deficiency heterozygotes or Canton-S wild-type flies. This result indicates that the DCortactin locus is deleted in the chromosome Df(3R)e-R1, but not in Df(3R)e-BS2, Df(3R)e-N19, or Df(3R)e-F1. Thus, the DCortactin gene was mapped on the chromosome segment between 93B3 and 93B7.

CONCLUSIONS

(1) We have isolated the gene encoding DCortactin. The gene consists of four exons. The protein coding sequence

Parsons, 1993). The isoforms are thought to be generated by alternative splicing. On the other hand, DCortactin has 4 cortactin repeats (Katsube et al., 1998). We found the third intron separating the third and fourth repeats coding sequences on the DCortactin gene. This exon-intron organization may be a molecular architecture for generating isoforms with different numbers of cortactin repeats. In fact, we identified a similar exon-intron organization of vertebrate cortactin genes (manuscript in preparation). To verify whether DCortactin has isoforms, we performed RT-PCR analysis designed to amplify the cortactin repeats coding region of the DCortactin transcript. Unexpectedly, only a product with the size corresponding to the transcript for 4 cortactin repeats was amplified from the RNA samples of embryo, larva, pupa, and adult fly (Fig. 4). Different from vertebrate cortactin, DCortactin is not likely to have isoforms which differ at the cortactin repeats region.

4. Cytological location of the DCortactin gene

The DCortactin gene was mapped cytologically to the 93B region on the right arm of the third chromosome (Fig. 5 A). To confirm the chromosomal position of the gene, genomic Southern blot analysis of flies heterozygous for some deficiency chromosomes was performed using a DCortactin cDNA probe (Fig. 5 B). A 4.2-kb band is about half the intensity in Df(3R)e-R1 heterozygotes compared to that in other deficiency heterozygotes or Canton-S wild-type flies. This result indicates that the DCortactin locus is deleted in the chromosome Df(3R)e-R1, but not in Df(3R)e-BS2, Df(3R)e-N19, or Df(3R)e-F1. Thus, the DCortactin gene was mapped on the chromosome segment between 93B3 and 93B7.

CONCLUSIONS

(1) We have isolated the gene encoding DCortactin. The gene consists of four exons. The protein coding sequence
is in the second, third, and fourth exons. The cortactin repeats coding region is divided by an intron between the third and fourth repeats.

(2) A 5' RACE as well as an S1 nuclease protection analysis revealed the heterogeneity of the 5'-untranslated region of the DCortactin transcript. In addition to the multiple transcription start sites, alternative splicing within the 5'-untranslated region was revealed.

(3) An RT-PCR analysis of the transcripts suggested that DCortactin has no isoforms having a different number of cortactin repeats in contrast to vertebrate homologs.

(4) The DCortactin gene was mapped between 93B3 and 93B7 on the right arm of the third chromosome. The isolation of mutant alleles of the DCortactin gene is currently in progress.

REFERENCES


