Dev Genes Evol
DOI 10.1007/s00427-006-0097-0

SEQUENCE CORNER

PDZ-domain-binding sites are common among cadherins

Fabio Demontis · Bianca Habermann · Christian Dahmann

Received: 30 January 2006 / Accepted: 7 June 2006

Abstract Cadherins are Ca^{2+}-dependent cell adhesion molecules that play fundamental roles in animal development and homeostasis. A number of cadherins contain conserved binding sites for catenins in their cytoplasmic region that are important for the adhesive function of these cadherins by mediating their interaction to the cytoskeleton. However, most cadherins lack apparent binding sites for catenins and their cytoplasmic interacting partners are mostly unknown. In this paper, we show, using bioinformatics, that a number of insect and vertebrate cadherins lacking catenin-binding sites contain conserved consensus sequences for C-terminal PSD-95/Discs-large/ZO-1 (PDZ)-domain-binding sites. This suggests that PDZ-domain-containing proteins are common cytoplasmic interacting partners for cadherins lacking catenin-binding sites.

Keywords PDZ domain · Cadherin · Cell adhesion

Communicated by B.G. Herrmann

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00427-006-0097-0 and is accessible to authorized users.

F. Demontis · C. Dahmann (✉)
Max Planck Institute of Molecular Cell Biology and Genetics,
Pfotenhauerstrasse 108,
01307 Dresden, Germany
e-mail: dahmann@mpi-cbg.de

B. Habermann
Scionics Computer Innovation,
Tatzberg 47-51,
01307 Dresden, Germany

Introduction

The interaction of cells is fundamental for the development and homeostasis of multicellular organisms. Cadherins are a family of transmembrane proteins that provide molecular interactions between adjacent plasma membranes and thereby mediate cohesion and signaling between cells (Gumbiner 2005). A classifying feature of cadherins is the presence of multiple approximately 110 amino acids long domains termed cadherin repeats in their extracellular regions. In general, cadherins provide molecular links between adjacent cells through binding of the extracellular cadherin repeats on adjacent plasma membranes and the interaction of their intracellular region with cytoskeleton-associated proteins or signaling pathways.

Cadherins have been grouped into six subfamilies based mainly on protein domain composition and phylogenetic analysis (Nollet et al. 2000). Type-I and type-II cadherins mainly localize to adherens junctions and mediate the cohesion between cells through binding with their intracellular regions to catenins. Desmogleins and desmocollins are components of desmosomes and mediate the cohesion between cells through interactions with catenins and, consequently, intermediate filaments. Common to these four subfamilies of cadherins is the presence of conserved catenin-binding sites in their cytoplasmic region that are important for their adhesive function. Members of the fifth subfamily, Flamingo cadherins, are important for establishing planar cell polarity and axon target selection. Finally, protocadherins, which constitute the largest subfamily of cadherins, are involved in processes including tissue morphogenesis and the generation of synaptic specificity (Frank and Kemler 2002). Members of the two latter subfamilies, as well as a number of cadherins that fall outside these six subfamilies, have no apparent catenin-binding sites (Nollet

et al. 2000). A few cadherins lacking catenin-binding sites have recently been shown to bind instead with their cytoplasmic regions to proteins unrelated to catenins. Cadherin 23 and Ret, for example, physically interact with harmonin and Shank3, respectively (Boeda et al. 2002; Schuetz et al. 2004; Siemens et al. 2002). However, for most of the cadherins that lack catenin-binding sites, the cytoplasmic binding partners are unknown and their identification is an important step in understanding the functions of these cadherins (Frank and Kemler 2002). Harmonin and Shank3 proteins contain PSD-95/Discs-large/ZO-1 (PDZ)-domains (Boeda et al. 2002; Schuetz et al. 2004; Siemens et al. 2002), protein interaction domains known to organize submembranous protein complexes (Sheng and Sala 2001). PDZ domains bind to internal peptides of proteins or, more commonly, to short C-terminal peptides that can be grouped according to different consensus sequences into PDZ-domain-binding sites of class I, II, or III (Sheng and Sala 2001). A few reports have recently predicted the presence of C-terminal and internal PDZ-domain-binding sites in additional cadherins including *Drosophila melanogaster* Cad99C (Schlichting et al. 2005), and the vertebrate cadherins Fat1 (Ponassi et al. 1999), Protocadherin LKC (Okazaki et al. 2002), Mu-protocadherin (Goldberg et al. 2000), and Protocadherin 15 (Adato et al. 2005). The binding of cadherins to PDZ-domain-containing proteins may be limited to these few cadherins or, alternatively, may be a more general property of cadherins. To distinguish between these possibilities, we have performed a comprehensive bioinformatics analysis of the presence of PDZ-domain-binding sites in human and *Drosophila melanogaster* cadherins. Our results suggest that PDZ-domain-containing proteins are common cytoplasmic interacting partners for cadherins lacking catenin-binding sites.

Materials and methods

Drosophila melanogaster cadherin sequences were retrieved from http://www.flybase.org on the basis of the accession numbers previously reported (Hill et al. 2001). Human cadherin sequences were in part retrieved from a previous report (Nollet et al. 2000) and also by searching the NCBI database for protein sequences corresponding to the search words "Homo sapiens AND cadherin" and "Homo sapiens AND protocadherin". To obtain cadherin-repeat-containing sequences not annotated as cadherins, we retrieved cadherin-repeat containing proteins using the Simple Modular Architecture Search Tool (SMART, Schultz et al. 1998).

All known *Drosophila* and human cadherin protein sequences, retrieved as described above, were analyzed manually for the presence of C-terminal peptides corresponding to consensus sequences of PDZ-domain-binding sites, as previously reported (Sheng and Sala 2001). To find insect and vertebrate orthologs to *Drosophila* and human cadherins, BLASTp searches against the non-redundant NCBI protein database and tBLASTn searches against EST and genomic sequences were performed. Due to the incompleteness of available genomic sequence information in some organisms and the apparent absence of orthologs in other organisms, orthologs for the human cadherins containing PDZ-domain-binding sites could not be identified in all species that we were analyzing. For example, we could only identify an ortholog to human Fat2 in mouse, but not in *Xenopus laevis* or *Danio rerio*. In Table 2 we included, if present, only orthologs from one representative species of mammalians, amphibians, and teleost fish, even though in most cases additional orthologs were identified.

The cytoplasmic regions of cadherins were predicted employing the Simple Modular Architecture Research Tool (SMART) (Schultz et al. 1998). Amino acid identity and similarity were calculated using the pairwise sequence alignment algorithm ClustalW v1.4, included in MacVector 7.2 (Accelrys).

Accession numbers of the insect and vertebrate sequences employed in this study can be found in the Electronic Supplementary Material.

Results and discussion

We analyzed all known *Drosophila melanogaster* and human cadherin sequences for the presence of C-terminal PDZ-domain-binding sites and found that a number of *Drosophila* and human cadherins contain C-terminal sequences that match the consensus sequences for PDZ-domain-binding sites. Internal PDZ-domain-binding sites are not mainly defined by sequence but rather by their structural features, impeding their identification by bioinformatics (Harris and Lim 2001). To begin to address the functional importance of the putative C-terminal PDZ-domain-binding sites, we analyzed their evolutionary conservation. The degree of conservation of a motif is thought to be a measure of the likelihood that the motif is of functional relevance (Gutman et al. 2005). Thus, highly conserved motifs are considered to be more likely to be of functional relevance than weakly conserved motifs.

To estimate the degree of conservation of the PDZ-domain-binding sites, we identified orthologs of *Drosophila melanogaster* cadherins containing putative PDZ-domain-binding sites in other insects and orthologs of human cadherins containing putative PDZ-domain-binding sites in other vertebrates. We then analyzed whether these orthologs contained C-terminal sequences that matched the consensus sequences for PDZ-domain-binding sites.

Orthologs of the α, β, and γ subfamilies of human protocadherins could, except in one case, not be identified, as these cadherins are more closely related to one another than to cadherins from other vertebrates (Wu 2005). Members of the α, β, and γ protocadherin subfamilies that did contain a putative C-terminal PDZ-domain-binding site are listed in Table S1. However, because orthologs of the α, β, and γ protocadherins could not be identified and, therefore, an evolutionary conservation of the PDZ-domain-binding sites could not be assessed, the functional relevance of these putative PDZ-domain-binding sites remains unclear.

Four (28%) of the predicted 14 *Drosophila melanogaster* cadherins that lack catenin-binding sites contained C-terminal PDZ-domain-binding sites conserved in other insects (Table 1). Likewise, 17 (47%) of the 36 human cadherins that lack catenin-binding sites contained C-terminal PDZ-domain-binding sites conserved in other vertebrates (Table 2). Orthologs displayed in most cases (20/21) putative C-terminal PDZ-domain-binding sites of the same class. Frequently, amino acids adjacent to the C-terminal amino acids matching the consensus sequences for PDZ-domain-binding sites were also conserved, consistent with them contributing to binding specificity, as has been previously suggested (Sheng and Sala 2001). The C-terminal 5 amino acids containing the putative PDZ-domain-binding sites were often more highly conserved than the remaining cytoplasmic region of the cadherin (Tables 1 and 2), suggesting a positive selection for the presence of the putative PDZ-domain-binding sites. The high sequence conservation across evolution of the putative PDZ-domain-binding sites suggests that the PDZ-domain-binding sites might have functional significance.

Additionally, we noticed that in some cases the putative PDZ-domain-binding sites were only present in unique splice variants (Table 2, Schuetz et al. 2004), indicating that alternative splicing may produce different forms of cadherins with distinct abilities to interact with PDZ-domain-containing proteins. Thus, this could allow, for example, for the tissue-specific formation of cadherin-PDZ-domain-containing protein complexes.

We also note that PDZ-domain-binding sites were not always conserved between orthologs from vertebrates and insects. For example, vertebrate Ret cadherin contains a C-terminal PDZ-domain-binding site; however, its *Drosophila melanogaster* ortholog does not. This indicates that in some instances, the ability of cadherins to bind to PDZ-domain-containing proteins might have been acquired during vertebrate evolution.

We did not identify conserved PDZ-domain-binding sites in cadherins containing catenin-binding sites, suggesting that the direct interaction with PDZ-domain-containing proteins is a unique property of cadherins that lack catenin-binding sites. Cadherins containing catenin-binding sites may, however, bind PDZ-domain-containing proteins indirectly, for example, through catenins (Perego et al. 2000).

Our results suggest that cadherins lacking catenin-binding sites commonly interact with PDZ-domain-containing proteins (Fig. 1). PDZ-domain-containing proteins may, there-

Table 1 *Drosophila melanogaster* cadherins containing conserved C-terminal PDZ-domain-binding sites

Cadherin	C-terminal sequence	Class	Identity + similarity (%)	
			Five C-terminal amino acids	Cytoplasmic region
Dm Cad88C	LARKLETTEL	I	Ag–Am 100%	31%+15%
Ag Cad88C	CARNLETTEL	I	Dm–Am 100%	39%+17%
Am Cad88C	LGKKIETTEL	I	Dm–Ag 100%	36%+15%
Dm Cad99C	RSEVETTTEL	I	Ag–Am 100%	28%+19%
Ag Cad99C	RSDAETTTEL	I	Dm–Am 100%	36%+17%
Am Cad99C	QSPVETTTEL	I	Dm–Ag 100%	46%+14%
Dm Cad96Ca	NIVSLSGEKL	III	Ag–Am 100%	60%+17%
Ag Cad96Ca	NMLNMSGEKL	III	Dm–Am 100%	54%+13%
Am Cad96Ca	NVLNLSGEKL	III	Dm–Ag 100%	68%+11%
Dm Fat	NGPAAPEEYV	III	Ag–Am 100%	55%+10%
Ag Fat	VQSKPSEEYV	III	Dm–Am 80%+0%	48%+15%
Am Fat	NIPKPSEEYV	III	Dm–Ag 80%+0%	58%+ 7%

The C-terminal ten amino acids are shown and the class of the PDZ-domain-binding site is indicated. Identity and similarity of the five C-terminal amino acids and the entire cytoplasmic region of orthologous cadherins are given as a percentage. The five C-terminal amino acids are more highly conserved than the overall cytoplamic region, indicating a positive selection for the putative PDZ-domain-binding sites. PDZ-domain-binding site consensus sequences are: class I (S/T, X, ϕ), class II (ϕ, X, ϕ), and class III (D/E, X, ϕ), with X=any amino acid and ϕ=hydrophobic amino acids A, I, L, M, F, or V. Ag–Am, Dm–Am, and Dm–Ag indicate sequence comparisons between the orthologous cadherins from the two respective species.

Dm *Drosophila melanogaster*, Ag *Anopheles gambiae*, Am *Apis mellifera*

Table 2 Human cadherins containing conserved C-terminal PDZ-domain-binding sites

Cadherin	C-terminal sequence	Class	Identity + similarity (%)	
			Five C-terminal amino acids	Cytoplasmic region
Hs Cadherin 23 is.1	METPLEITEL	I	Xt–Hs 100%	65%+ 9%
Xt Cadherin 23	EEGPLEITEL	I	Xt–Dr 100%	63%+ 9%
Dr Cadherin 23 is.1	SESPLEITEL	I	Hs–Dr 100%	74%+14%
Hs Fat 1	PLDSQQHTEV	I	Xt–Hs 80%+20%	74%+12%
Xt Fat 1	PLDSQHHTEV	I	Xt–Dr 80%+20%	64%+12%
Dr Fat 1	DSQHQQHTEV	I	Hs–Dr 100%	64%+14%
Hs Fat3	FVETQHQTQV	I	Xt–Hs 100%	68%+ 8%
Xt Fat3	YMETQHQTQV	I	Xt–Fr 60%+40%	45%+12%
Fr Fat3	ITDSEQQTKV	I	Hs–Fr 60%+40%	45%+10%
Hs Protocadherin LKC	TNAGLDTTDL	I	Xt–Hs 100%	39%+12%
Xt Protocadherin LKC	SNGALDTTDL	I	Xt–Tn 80%+ 0%	30%+16%
Tn Protocadherin LKC	NNPAFSTTDL	I	Hs–Tn 80%+ 0%	37%+14%
Hs Mu-protocadherin is.1, is.3	DAPGGDDSYI	I	Xl–Hs 20%+20%	23%+13%
Xl Mu-protocadherin	NNYSDNFSQL	I		
Hs Ret isoform c	GRISHAFTRF	I	Xl–Hs 100%	94%+ 4%
Xl Ret	GRISHAFTRF	I	Xl–Dr 100%	82%+ 2%
Dr Ret	GRISHAFTRF	I	Hs–Dr 100%	77%+ 7%
Hs Protocadherin 15	KQSHSQSTSL	I	Xt–Hs 80%+20%	18%+ 9%
Xt Protocadherin 15	KRFPSQSTAL	I	Xt–Dr 80%+ 0%	30%+ 7%
Dr Protocadherin 15b	KRFPSQSTVL	I	Hs–Dr 80%+ 0%	29%+13%
Hs Protocadherin 20	ERKPMDISNI	I	Xt–Hs 80%+20%	60%+10%
Xt Protocadherin 20	EKRPIEISNI	I	Xt–Fr 80%+20%	26%+31%
Fr Protocadherin 20	NRKLRQISNI	I	Hs–Fr 80%+ 0%	36%+13%
Hs Cadherin 16	ADSVPLKATV	II	Xl–Hs 20%+20%	9%+23%
Xl Cadherin 16	SKKSRKSDAV	III	Xl–Dr 0%+20%	11%+16%
Dr Cadherin 16	NKDCQERVPL	II	Hs–Dr 0%+40%	19%+14%
Hs Dachsous 1	EPPDDTELHI	II	Xt–Hs 60%+20%	57%+13%
Xt Dachsous 1	DPVGDGELRI	II	Xt–Fr 80%+ 0%	63%+15%
Fr Dachsous 1	EHIEEAELRI	II	Hs–Fr 60%+40%	60%+10%
Hs Dachsous 2 is.1	ELKAEDEVQI	II	Xt–Hs 40%+40%	46%+18%
Xt Dachsous 2	ELLIQQEFQV	II	Xt–Tn 60%+ 0%	34%+14%
Tn Dachsous 2	ASLLEAEIQV	II	Hs–Tn 40%+20%	30%+16%
Hs Fat 2	DYGSCEEVMF	II	Mm–Hs 100%	83%+ 5%
Mm Fat 2	DYGSCEEVMF	II		
Hs Protocadherin 1 is.2	AQTAKREIYL	II	Xt–Hs 100%	55%+ 7%
Xt Protocadherin 1	LSSAKREIYL	II	Xt–Tn 100%	38%+ 8%
Tn Protocadherin 1 is.2	VHTPKREIYL	II	Hs–Tn 100%	55%+10%
Hs Protocadherin 7 is.c	NTLTRREVYL	II	Xt–Hs 100%	61%+ 6%
Xt Protocadherin 7	NTLTRREVYL	II	Xt–Tn 100%	29%+10%
Tn Protocadherin 7	LTLSRREVYL	II	Hs–Tn 100%	41%+13%
Hs Protocadherin 19	GVKRLKDIVL	II	Xt–Hs 100%	79%+ 9%
Xt Protocadherin 19	GVKRLKDIVL	II	Xt–Dr 100%	65%+10%
Dr Protocadherin 19	GSKRLKDIVL	II	Hs–Dr 100%	63%+11%
Hs Protocadherin 8 is.1, is.2	SPKKGANENV	III	Xl–Hs 0%+60%	52%+ 7%
Mm Protocadherin 8	SPKKGINENV	III	Xl–Mm 0%+60%	55%+11%
Xl Protocadherin 8	DDSTCEQDEL	III	Hs–Mm 80%+ 0%	64%+ 1%
Hs Protocadherin β 1	GHDQVSDDYM	III	Rn–Hs 80%+ 0%	78%+ 8%
Mm Protocadherin β 1	GHDRIGDEYM	III	Mm–Hs 60%+20%	77%+12%
Rn Protocadherin β 1	GHARAGDDYM	III	Mm–Rn 80%+20%	89%+ 4%

The ten C-terminal amino acids are shown and the class of the PDZ-domain-binding site is indicated. PDZ-domain-binding-site consensus sequences are as shown in Table 1. Identity and similarity of the five C-terminal amino acids and the entire cytoplasmic region of orthologous cadherins are given as a percentage. The five C-terminal amino acids are frequently more highly conserved than the overall cytoplasmic region, indicating a positive selection for the putative PDZ-domain-binding sites. Xt-Hs, Xt-Dr, Hs-Dr, Xt-Fr, Hs-Fr, Xt-Tn, Hs-Tn, Xl-Hs, Xl-Dr, Mm-Hs, Xl-Mm, Hs-Mm, Rn-Hs, and Mm-Rn indicate sequence comparisons between the orthologous cadherins from the two respective species.
Hs *Homo sapiens*, Mm *Mus musculus*, Rn *Rattus norvegicus*, Xt *Xenopus tropicalis*, Xl *Xenopus laevis*, Dr *Danio rerio*, Tn *Tetraodon nigroviridis*, Fr *Fugu rubripes*, *is.* isoform

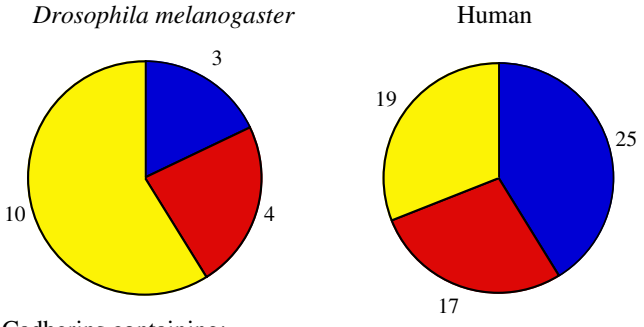

Fig. 1 PDZ-domain-binding sites are a common motif of insect and human cadherins not containing catenin-binding sites. Pie-chart representations of the number of cadherins in *Drosophila melanogaster* and humans containing conserved consensus sequences for C-terminal PDZ-domain-binding sites or catenin-binding sites. α, β, and γ subfamilies of human protocadherins are not included (see text)

fore, be important effector molecules mediating and regulating the functions of this group of cadherins. The binding of PDZ-domain-containing proteins to cadherins lacking a catenin-binding site may provide these cadherins with a linkage to the actin cytoskeleton (Boeda et al. 2002; Siemens et al. 2002), modulate their adhesive properties, or contribute to their signalling functions either directly (Okazaki et al. 2002) or indirectly by assembling supramolecular complexes (Sheng and Sala 2001), similar to the way catenins regulate the function of cadherins containing catenin-binding sites.

Acknowledgements We thank Karla Neugebauer and Heather Thompson for critical comments on the manuscript. This work was supported by the Max Planck Society.

References

Adato A, Michel V, Kikkawa Y, Reiners J, Alagramam KN, Weil D, Yonekawa H, Wolfrum U, El-Amraoui A, Petit C (2005) Interactions in the network of Usher syndrome type 1 proteins. Hum Mol Genet 14:347–356

Boeda B, El-Amraoui A, Bahloul A, Goodyear R, Daviet L, Blanchard S, Perfettini I, Fath KR, Shorte S, Reiners J, Houdusse A, Legrain P, Wolfrum U, Richardson G, Petit C (2002) Myosin VIIa, harmonin and cadherin 23, three Usher I gene products that cooperate to shape the sensory hair cell bundle. EMBO J 21:6689–6699

Frank M, Kemler R (2002) Protocadherins. Curr Opin Cell Biol 14:557–562

Goldberg M, Peshkovsky C, Shifteh A, Al-Awqati Q (2000) Mu-Protocadherin, a novel developmentally regulated protocadherin with mucin-like domains. J Biol Chem 275:24622–24699

Gumbiner BM (2005) Regulation of cadherin-mediated adhesion in morphogenesis. Nat Rev Mol Cell Biol 6:622–634

Gutman R, Berezin C, Wollman R, Rosenberg Y, Ben-Tal N (2005) QuasiMotiFinder: protein annotation by searching for evolutionarily conserved motif-like patterns. Nucleic Acids Res 33: W255–W261

Harris BZ, Lim WA (2001) Mechanism and role of PDZ domains in signaling complex assembly. J Cell Sci 114:3219–3231

Hill E, Broadbent ID, Chothia C, Pettitt J (2001) Cadherin superfamily proteins in *Caenorhabditis elegans* and *Drosophila melanogaster*. J Mol Biol 305:1011–1024

Nollet F, Kools P, van Roy F (2000) Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members. J Mol Biol 299:551–572

Okazaki N, Takahashi N, Kojima S, Masuho Y, Koga H (2002) Protocadherin LKC, a new candidate for a tumor suppressor of colon and liver cancers, its association with contact inhibition of cell proliferation. Carcinogenesis 23:1139–1148

Perego C, Vanoni C, Massari S, Longhi R, Pietrini G (2000) Mammalian LIN-7 PDZ proteins associate with beta-catenin at the cell–cell junctions of epithelia and neurons. EMBO J 19: 3978–3989

Ponassi M, Jacques TS, Ciani L, Ffrench Constant C (1999) Expression of the rat homologue of the *Drosophila* fat tumour suppressor gene. Mech Dev 80:207–212

Schlichting K, Demontis F, Dahmann C (2005) Cadherin Cad99C is regulated by Hedgehog signaling in *Drosophila*. Dev Biol 279:142–154

Schuetz G, Rosario M, Grimm J, Boeckers TM, Gundelfinger ED, Birchmeier W (2004) The neuronal scaffold protein Shank3 mediates signaling and biological function of the receptor tyrosine kinase Ret in epithelial cells. J Cell Biol 167:945–952

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci USA 95:5857–5864

Sheng M, Sala C (2001) PDZ domains and the organization of supramolecular complexes. Annu Rev Neurosci 24:1–29

Siemens J, Kazmierczak P, Reynolds A, Sticker M, Littlewood-Evans A, Muller U (2002) The Usher syndrome proteins cadherin 23 and harmonin form a complex by means of PDZ-domain interactions. Proc Natl Acad Sci USA 99:14946–14951

Wu Q (2005) Comparative genomics and diversifying selection of the clustered vertebrate protocadherin genes. Genetics 169: 2179–2188