Experimental Cell Research 358 (2017) 3–9

Contents lists available at ScienceDirect

Experimental Cell Research

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

Evolution and diversity of cadherins and catenins

Ismail Sahin Gul[a,b,1], Paco Hulpiau[a,b,1], Yvan Saeys[a,c], Frans van Roy[a,b,*]

[a] Center for Inflammation Research, VIB, Ghent, Belgium
[b] Department of Biomedical Molecular Biology, Ghent University, Ghent, Belgium
[c] Department of Internal Medicine, Ghent University, Ghent, Belgium

ARTICLE INFO

Keywords:
Cadherin superfamily
Armadillo superfamily
Catenins
Molecular evolution
Cell adhesion
Protocadherins
α-catenin

ABSTRACT

Cadherin genes encode a superfamily of conserved transmembrane proteins that share an adhesive ectodomain composed of tandem cadherin repeats. More than 100 human cadherin superfamily members have been identified, which can be classified into three families: major cadherins, protocadherins and cadherin-related proteins. These superfamily members are involved in diverse fundamental cellular processes including cell-cell adhesion, morphogenesis, cell recognition and signaling. Epithelial cadherin (E-cadherin) is the founding cadherin family member. Its cytoplasmic tail interacts with the armadillo catenins, p120 and β-catenin. Further, α-catenin links the cadherin/armadillo catenin complex to the actin filament network. Even genomes of ancestral metazoan species such as cnidarians and placozoans encode a limited number of distinct cadherins and catenins, emphasizing the conservation and functional importance of these gene families. Moreover, a large expansion of the cadherin and catenin families coincides with the emergence of vertebrates and reflects a major functional diversification in higher metazoans. Here, we revisit and review the functions, phylogenetic classifications and co-evolution of the cadherin and catenin protein families.

1. Metazoan evolution and diversity of cadherins

1.1. Classification and adhesion modes of cadherin superfamily members

The cadherin superfamily comprises calcium-dependent membrane proteins involved in cell-cell adhesion and cell-cell recognition. Each protein possesses at least two consecutive extracellular cadherin (EC) repeats and belongs to one of three families: the cadherin (CDH) family proper, the protocadherin (PCDH) family or the cadherin-related (CDHR) family [1]. The human reference genome encodes 114 protein encoding genes, which can be classified into one of these three families on the basis of their evolutionary history and their functional and structural features. Considering their respective sequence homologies each family can be further subdivided into smaller subfamilies with more closely related members (Table 1).

The founding member of the superfamily is E-cadherin or cadherin 1 (CDH1), which is one of the 32 members of the major cadherin family found in humans. Structural studies showed a similar interaction mode in all type-I classical cadherins, in cadherin 26 (CDH26) and in desmosomal cadherins (Fig. 1A). The N-terminal cadherin repeat (EC1) forms a homophilic interaction in *trans* with the same cadherin type on the opposing cell surface. A conserved tryptophan residue (Trp2) in the so-called adhesion arm is inserted in the hydrophobic pocket of the adhesion partner forming a strand-swap dimer [2]. The second cadherin repeat (EC2) interacts in *cis* with the EC1 of the neighboring cadherin on the same cell surface. Type-II classical cadherins use the same homophilic interaction mechanism but here two tryptophan residues (Trp2 and Trp4) are inserted into a larger hydrophobic pocket [3]. 7D cadherins and the CELSR cadherins have more than five EC repeats [1]. Their exact interaction mechanism is currently unknown. CELSR cadherins do not have a conserved Trp like the other members of the cadherin family and will therefore use another interaction mechanism. Interestingly, the 7D cadherin CDH17 and E-cadherin (CDH1) can form heterophilic *trans* interactions with each other [4]. 7D cadherins contain seven EC repeats and have one conserved Trp in their third EC repeat (EC3), which could form the typical strand-swap dimer with the EC1 from classical cadherins having five EC repeats. Except for these 7D cadherins, the three CELSR cadherins and CDH13, all other members in the major cadherin family have conserved cytoplasmic domains, which can interact with armadillo catenins as described in some detail below.

* Correspondence to: Department of Biomedical Molecular Biology & Center for Inflammation Research, Ghent University & VIB, FSVM Building, Technologiepark 927, B-9052 Ghent, Belgium.
E-mail address: frans.vanroy@ugent.be (F. van Roy).
[1] Contributed equally.

http://dx.doi.org/10.1016/j.yexcr.2017.03.001
Received 16 January 2017; Received in revised form 27 February 2017; Accepted 1 March 2017
Available online 06 March 2017

Table 1
Classification and official nomenclature of the 114 human cadherin protein encoding genes.

Family	Subfamily	Official gene names (HGNC symbol, common alias)
Major cadherins (CDH) *32 members*	Type I classical cadherins	cadherin 1 (**CDH1**, E-cadherin), cadherin 2 (**CDH2**, N-cadherin), cadherin 3 (**CDH3**, P-cadherin), cadherin 4 (**CDH4**, R-cadherin), cadherin 15 (**CDH15**, M-cadherin)
	Type II classical cadherins	cadherin 5 (**CDH5**, VE-cadherin), cadherin 6 (**CDH6**, K-cadherin), cadherin 7 (**CDH7**), cadherin 8 (**CDH8**), cadherin 9 (**CDH9**, T1-cadherin), cadherin 10 (**CDH10**, T2-cadherin), cadherin 11 (**CDH11**, OB-cadherin), cadherin 12 (**CDH12**, N-cadherin 2), cadherin 18 (**CDH18**), cadherin 19 (**CDH19**), cadherin 20 (**CDH20**), cadherin 22 (**CDH22**), cadherin 24 (**CDH24**)
	7D cadherins	cadherin 16 (**CDH16**, Ksp-cadherin), cadherin 17 (**CDH17**, LI-cadherin)
	Desmosomal cadherins	desmocollin 1 (**DSC1**), desmocollin 2 (**DSC2**), desmocollin 3 (**DSC3**), desmoglein 1 (**DSG1**), desmoglein 2 (**DSG2**), desmoglein 3 (**DSG3**), desmoglein 4 (**DSG4**)
	Flamingo or CELSR cadherins	cadherin EGF LAG seven-pass G-type receptor 1 (**CELSR1**), cadherin EGF LAG seven-pass G-type receptor 2 (**CELSR2**), cadherin EGF LAG seven-pass G-type receptor 3 (**CELSR3**)
	–	cadherin 13 (**CDH13**, H-cadherin), cadherin 26 (**CDH26**)
Protocadherins (PCDH) *65 members*	Clustered protocadherins	protocadherin alpha cluster (PCDHA@: **PCDHAC1**, **PCDHAC2**, **PCDHA1 up to PCDHA14**), protocadherin beta cluster (PCDHB@: **PCDHB1 up to PCDHB16**), protocadherin gamma cluster (PCDHG@: **PCDHGA1 up to PCDHGA12**, **PCDHGB1 up to PCDHGB7**, **PCDHGC3**, **PCDHGC4**, **PCDHGC5**)
	Non-clustered protocadherins	protocadherin 1 (**PCDH1**), protocadherin 7 (**PCDH7**), protocadherin 8 (**PCDH8**), protocadherin 9 (**PCDH9**), protocadherin 10 (**PCDH10**), protocadherin 11 X-linked and Y-linked (**PCDH11X** and **PCDH11Y**), protocadherin 12 (**PCDH12**), protocadherin 17 (**PCDH17**), protocadherin 18 (**PCDH18**), protocadherin 19 (**PCDH19**), protocadherin 20 (**PCDH20**)
Cadherin-related (CDHR) *17 members*	–	cadherin related family member 1 (**CDHR1**, PCDH21), cadherin related family member 2 (**CDHR2**, PCDH24), cadherin related family member 3 (**CDHR3**, CDH28), cadherin related family member 4 (**CDHR4**, CDH29), cadherin related family member 5 (**CDHR5**, MU-PCDH), protocadherin related 15 (**PCDH15**, CDHR15), ret proto-oncogene (**RET**, CDHR16), cadherin related 23 (**CDH23**, CDHR23)
	Dachsous	dachsous cadherin-related 1 (**DCHS1**, CDHR6), dachsous cadherin-related 2 (**DCHS2**, CDHR7)
	FAT	FAT atypical cadherin 1 (**FAT1**, CDHR8), FAT atypical cadherin 2 (**FAT2**, CDHR9), FAT atypical cadherin 3 (**FAT3**, CDHR10), FAT atypical cadherin 4 (**FAT4**, CDHR11)
	Calsyntenins	calsyntenin 1 (**CLSTN1**, CDHR12), calsyntenin 2 (**CLSTN2**, CDHR13), calsyntenin 3 (**CLSTN3**, CDHR14)

Official human gene nomenclature (HGNC) symbols are in bold and underlined. Human calcium-dependent membrane proteins with at least two consecutive extracellular cadherin (EC) repeats can be classified into three large families (first column) based on phylogenetic analysis Ref. [1]. The family subtrees in such evolutionary analysis reveal several subfamilies (second column). Members of the subfamilies (third column) have similar functional and structural features as described in the main text.

Protocadherins (PCDH) form in vertebrates the largest cadherin family composed of two subfamilies: clustered and non-clustered protocadherins (Table 1). They play a crucial role in the vertebrate nervous system by generating cell surface diversity and specificity, in that way determining the cellular identity of individual neurons [5]. Mammals have three clusters of protocadherin genes, α-PCDH, β-PCDH and γ-PCDH, organized consecutively on the genome. At the cell surface promiscuous *cis* dimer formation occurs between the membrane-proximal EC repeats (EC6) (Fig. 1B). Two models have been proposed for explaining how such dimeric recognition units can engage in *trans* interface binding [6]. Tetramers can be formed by head-to-tail EC1 to EC4 *trans* interactions of the *cis* dimers. In a second model *cis* dimers bind two dimeric recognition units on the opposing cell surface and by repeating this a zipper-like assembly can be established (Fig. 1B). The genes of the non-clustered protocadherin subfamily, also called delta-protocadherins (δ-PCDH), are largely dispersed in the genome. Part of the δ-PCDH proteins (the δ2-PCDHs) have six EC repeats like the clustered protocadherins, others (the δ1-PCDHs) possess seven EC repeats. The structure of zebrafish PCDH19, a δ2-PCDH, revealed a “forearm handshake” adhesion involving EC1 tot EC4 domains resulting in a fully overlapping antiparallel, homophilic *trans* dimer (Fig. 1C, [7]). This binding mechanism is similar to the *trans* interaction mode of clustered protocadherins. Other non-clustered protocadherins are expected to use the same adhesive interface.

The cadherin-related family (CDHR) gathers the most diverse members. The calsyntenins, involved in learning and memory formation, have only two EC repeats, which are together able to mediate either homophilic or heterophilic adhesive interactions [8]. It is currently unclear if their primary activity is cell-cell adhesion. Other functions such as secreted ligands or transport chaperones have been suggested. The Dachsous (DCHS1, DCHS2) and FAT (FAT1, −2,−3 and FAT4) cadherins are among the longest cadherins. Heterophilic interactions between DCHS1 with 27 EC repeats and FAT4 with 34 EC repeats regulate planar cell polarity and cell proliferation [9]. Classical cadherins are typically found in adherens junctions (AJs). To fit in a 30- to 45-nm intercellular junction, only about double the size of AJs, the giant cadherins were shown to self-bend at certain EC-EC linkers where the typical calcium binding amino acids (AAs) are not conserved [9] (Fig. 1D). Two other cadherin-related proteins PCDH15 (CDHR15) and CDH23 (CDHR23) are found at tip links of stereocilia in the inner ear and are part of the hair-cell transduction machinery [10]. Note that although their official gene symbols refer to protocadherins and classical cadherins, PCDH15 (CDHR15) with 11 EC repeats and CDH23 (CDHR23) with 27 EC repeats are not a protocadherin and not a cadherin, respectively [1]. They form specific heterophilic *trans* interaction complexes by means of an extended handshake of the EC1-EC2 repeats between a PCDH15 *cis* dimer on the one hand and a CDH23 *cis* dimer on the other hand (Fig. 1E). Lastly, CDHR2 (PCDH24) and CDHR5 (MU-PCDH) can also heterophilically *trans*-interact to form intermicrovillar adhesion links between adjacent microvilli in the intestinal brush border [11]. Like PCDH15 and CDH23, CDHR2 with nine EC repeats and CDHR5 with four EC repeats have different domain compositions and phylogenetic positions in the superfamily than the protocadherins, but both are often still incorrectly called a protocadherin (PCDH24 and MU-PCDH) rather than a cadherin-related protein.

1.2. Evolution of the cadherin superfamily

The first proteins able to mediate calcium-dependent cell-cell adhesion by means of at least two EC repeats and by this definition belonging to the cadherin superfamily, appeared in the last common ancestor of animals more than 600 million years ago. Only a few