

Functional implications of γ -Protocadherin structural diversity

Kerry Marie Goodman^{1,6}, Rotem Rubinstein^{1,2,6}, Chan Aye Thu¹,

Seetha Mannepalli¹, Fabiana Bahna^{1,3}, Göran Ahlsén^{2,3}, Chelsea Rittenhouse¹,

Tom Maniatis^{1,5}, Barry Honig^{1,2,3,4,5*}, and Lawrence Shapiro^{1,2,5*}

¹Department of Biochemistry and Molecular Biophysics, ²Department of Systems Biology, ³Howard Hughes Medical Institute, ⁴Department of Medicine, and ⁵Zuckerman Mind Brain and Behavior Institute, Columbia University, New York, NY 10032, USA.

⁶ Equal contribution

* Correspondence: bh6@cumc.columbia.edu (B.H.); shapiro@convex.hhmi.columbia.edu (L.S.)

Abstract

Stochastic cell-surface expression of α -, β -, and γ -clustered protocadherins (Pcdhs) provides vertebrate neurons with single-cell identities that underlie neuronal self-recognition. Here we report crystal structures of ectodomain fragments from *trans*-interacting regions of γ -Pcdhs γ A1, γ A8, and γ B7 forming homodimers and of C-terminal ectodomain fragments from γ -Pcdhs γ A4 and γ B2, which depict *cis*-interacting regions in monomeric form, providing structures that span the entire Pcdh ectodomain. The *trans*-dimer structures reveal determinants of γ -Pcdh isoform-specific homophilic recognition. We identified and structurally mapped *cis*-dimerization mutations to the C-terminal ectodomain structures. Biophysical studies showed that Pcdh ectodomains from γ B-subfamily isoforms formed *cis* dimers, whereas γ A isoforms did not, but both γ A and γ B isoforms could interact in *cis* with α -Pcdhs. Together, these data show how interaction specificity is distributed over all domains of the γ -Pcdh *trans* interface, and suggest that subfamily- or isoform-specific *cis*-interactions could play a role in the Pcdh-mediated neuronal self-recognition code.

Introduction

A characteristic of neural circuit assembly is that dendrites and axonal arbors of the same neuron do not stably contact one another, but are free to interact with the processes of other neurons (Zipursky and Grueber, 2013; Zipursky and Sanes, 2010). This fundamental property of neural circuit assembly is accomplished through a mechanism that mediates ‘self-avoidance’ between sister branches from individual neurons, while permitting interactions between non-self neurons. In both vertebrates and invertebrates, self-avoidance is thought to rely on the generation of unique single cell surface identities through mechanisms that involve the stochastic expression of unique combinations of cell surface protein isoforms (Zipursky and Grueber, 2013; Chen and Maniatis, 2013). In *Drosophila* and many other invertebrates individual-neuron identities are provided by the expression of single-cell-specific Dscam1-isoform subsets generated by stochastic alternative splicing (Miura et al., 2013; Schmucker et al., 2000; Wojtowicz et al., 2004; Neves et al., 2004; Zhan et al., 2004). By contrast, in vertebrates the clustered protocadherins (Pcdhs) provide analogous cell-surface diversity, but in this case generated through stochastic alternative promoter choice (Tasic et al., 2002; Wang et al., 2002; Esumi et al., 2005; Hirano et al., 2012; Kaneko et al., 2006)..

Both the invertebrate Dscam1 proteins, and vertebrate Pcdhs are highly diverse families of cell-surface proteins that form isoform-specific trans-dimers between apposed neuronal cell surfaces (Zipursky and Grueber, 2013; Zipursky and Sanes, 2010; Chen and Maniatis, 2013; Thu et al., 2014; Schreiner and Weiner, 2010). Stochastic alternative splicing of the *Dscam1* gene in *D. melanogaster* produces up to 19,008 distinct protein isoforms, the majority of which engage in highly specific *trans* homodimerization (Miura

et al., 2013; Schmucker et al., 2000; Wojtowicz et al., 2004, 2007). In contrast, mice and humans express just 58 and 53 distinct Pcdh isoforms, respectively, each of which display isoform-specific homophilic binding in *trans* (Schreiner and Weiner, 2010; Thu et al., 2014). Biophysical measurements with domain-deleted proteins showed that Pcdhs also interact in *cis*, through a membrane-proximal dimer interface involving extracellular cadherin domain 6 (EC6) and potentially EC5 (Thu et al., 2014; Rubinstein et al., 2015). Pcdh *cis* dimers are thought to form promiscuously (Schreiner and Weiner, 2010; Thu et al., 2014), and thus provide a large repertoire of *cis* dimeric Pcdh recognition units (Rubinstein et al., 2015; Thu et al., 2014).

Vertebrate protocadherin genes have a unique organization in which the α , β , and γ gene clusters are arranged in tandem (Wu and Maniatis, 1999). Each of the Pcdh gene clusters contains multiple alternative variable exons (14 α , 22 β , and 22 γ in the mouse) which encode full Pcdh ectodomains, including six extracellular cadherin (EC) domains, a single transmembrane region and a short cytoplasmic extension. The *Pcdh α* and *Pcdh γ* gene clusters also contain three ‘constant’ exons that encode cluster-specific intracellular domains. The two variable exons in the *Pcdh α* gene cluster and the last three variable exons of the *Pcdh γ* gene cluster are divergent from other Pcdh ‘alternate’ isoforms and are referred to as ‘C-type’ Pcdhs (Wu and Maniatis, 1999; Wu et al., 2001). The non-C-type *Pcdh γ* genes have been further divided into two subfamilies—*Pcdh γ A* and *Pcdh γ B*—based on sequence identity/phylogenetic analysis (Wu and Maniatis, 1999). Single-cell RT-PCR studies of the *Pcdh α* and *Pcdh γ* clusters in Purkinje neurons revealed that each neuron expresses all C-type Pcdhs biallelically, along with ~10 alternate isoforms (α , β &

γ) stochastically expressed from each gene cluster independently on allelic chromosomes (Esumi et al., 2005; Kaneko et al., 2006).

Each of the three Pcdh families may serve specialized functions. *Pcdh α* knockouts of individual gene clusters revealed neuronal wiring defects in olfactory and serotonergic neurons (Hasegawa et al., 2008, 2012; Katori et al., 2009). By contrast, genetic ablation of the *Pcdh γ* gene cluster leads to lethality at P0 (Lefebvre et al., 2008; Wang et al., 2002), and revealed a cell-death phenotype for some neuron types (Wang et al., 2002; Weiner et al., 2005; Lefebvre et al., 2008; Prasad and Weiner, 2011; Chen et al., 2012). Conditional deletion of the *Pcdh γ* cluster which bypasses neonatal lethality, revealed defects in dendritic arborization of cortical neurons (Garrett et al., 2012). Similarly, γ -Pcdh knockdown in hippocampal neurons in vitro resulted in dendritic arbors with lower complexity (Suo et al., 2012). Subsequent studies with transgenic and conditional knockout mice suggest that γ -Pcdhs act locally to regulate dendrite arborization, with the complexity of a neuron's dendritic arbor determined, at least in part, by Pcdh-dependent non-cell autonomous interaction of a neuron with surrounding neurons and glia (Molumby et al., 2016).

The clustered Pcdhs were first implicated in dendritic self-avoidance through studies of the *Pcdh γ* -gene cluster. Deletion of all 22 genes of the *Pcdh γ* cluster in mice results in a loss of dendritic self-avoidance in retinal starburst amacrine cells (SACs) and cerebellar Purkinje cells (Lefebvre et al., 2012), with formation of self-synapses (autapses) in SACs (Kostadinov and Sanes, 2015). However, most other neuron types appeared unaffected by the loss of the *Pcdh γ* gene cluster.

Cellular recognition specificities of Pcdhs appear to be diversified by co-expression of multiple Pcdh isoforms in the same cell (Yagi et al., 2013; Schreiner and Weiner, 2010; Thu et al., 2014). In general, recognition between cells expressing multiple Pcdhs is only observed when all expressed isoforms match. In early work, Schreiner and Weiner (2010) showed that expression of mismatched isoforms resulted in less binding between a cell population adhered to a surface and cells passed over them. We assessed the ability of cells co-transfected with up to five Pcdh isoforms to co-aggregate with cells containing various numbers of mismatches, and found that expression of even a single mismatch prevented co-aggregation in cell aggregation assays (Thu et al., 2014). Thus, even a single mismatched isoform is able to interfere with recognition. Importantly, this behavior – which we termed ‘interference’ – is not observed with classical cadherins (Thu et al., 2014). We therefore suggested that the interference phenomenon could arise from promiscuous *cis* dimerization between co-expressed Pcdh isoforms to form single-cell repertoires of dimeric Pcdh recognition units (Rubinstein et al., 2015).

The specificity-determining cell-cell recognition interface of Pcdhs involves domains EC1–4, as shown experimentally through mutagenesis analysis (Rubinstein et al., 2015) and suggested by mutation correlation analysis (Nicoludis et al., 2015). Structures of the *trans* dimer formed through this interface have been reported for two α -Pcdhs and two β -Pcdhs (Goodman et al., 2016) and subsequently for a γ B-Pcdh (Nicoludis et al., 2016). As expected, all isoforms had overall-similar recognition-dimer structures, mediated by interfaces populated with diverse residue compositions that determine homophilic specificity (Goodman et al., 2016; Nicoludis et al., 2016). Here we report structures of recognition dimers from three γ -Pcdhs, two from the γ A subfamily, and one from the γ B

subfamily. The large collection of clustered Pcdh protein structures now available, including the first structures determined of *trans*-dimer engaged γ A Pcdhs, allowed us to analyze the specificity determinants across the clustered Pcdh family.

In addition to new *trans*-dimeric structures, we also present the first Pcdh structures that include the promiscuous *cis*-dimerization region, although in monomeric form.

Mutagenesis studies identify residues important for *cis* association and allow the visualization of these residues in the context of the structure. Finally, we show that Pcdh isoforms of the γ A and γ B Pcdh subfamilies differ in their *cis* associations, and we report variability among homophilic *cis* associations of C-type Pcdhs. These differences, along with those previously characterized for α -Pcdhs (Thu et al., 2014), suggests that individual isoform- or subfamily-differences in *cis* interaction behavior may play an important in generating a Pcdh self-recognition code.

Results

Trans interactions and Pcdh specificity

Crystal structures of γ -Pcdh cell-cell recognition dimers

To characterize the cell-cell recognition (*trans*) interfaces of γ -Pcdhs we produced EC1–4 or EC1–5 fragments of Pcdh γ A, γ B and γ C isoforms using suspension HEK293 cells.

These constructs encompassed the entire Pcdh EC1–4-mediated *trans* interface (Rubinstein et al., 2015; Nicoludis et al., 2015; Goodman et al., 2016), but lacked EC6, which mediates a distinct *cis* interface (Thu et al., 2014; Rubinstein et al., 2015). We used sedimentation equilibrium analytical ultracentrifugation (AUC) to characterize the homophilic binding properties of these proteins. The γ A isoform constructs— γ A1_{EC1–4}, γ A4_{EC1–4}, γ A8_{EC1–4}, γ A9_{EC1–5}—displayed dimer dissociation constants (K_{DS}) of between 8.6 and 45.3 μ M (Table 1). The γ B isoforms (γ B5_{EC1–4}, γ B6_{EC1–4}, γ B7_{EC1–4}) *trans* dimer affinities were more varied, with K_{DS} between 29 and 147 μ M (Table 1 and Rubinstein et al., 2015). Finally, both γ C isoform *trans*-interacting fragments tested— γ C3_{EC1–4} and γ C5_{EC1–5}—formed relatively weak dimers, with K_{DS} of 115 and 100 μ M respectively (Table 1 and Rubinstein et al., 2015).

Crystallization screening of these dimeric γ -Pcdh fragments yielded crystals of γ A1_{EC1–4}, γ A8_{EC1–4}, γ A9_{EC1–5}, and γ B7_{EC1–4}, and their structures were determined by molecular replacement (Figure 1A and Figure 1—figure supplement 1A). X-ray diffraction by the γ A9_{EC1–5}, and γ B7_{EC1–4} crystal form 1 crystals was significantly anisotropic and therefore the data was truncated using ellipsoidal limits for structure determination and refinement (Figure 1—source data 1 and Figure 1—figure supplement 2). The resolution of the

structures was 4.2 Å for $\gamma A1_{EC1-4}$, 3.6 Å for $\gamma A8_{EC1-4}$, 2.9/4.3/3.2 Å for $\gamma A9_{EC1-5}$, 4.5/4.5/3.6 Å for $\gamma B7_{EC1-4}$ crystal form 1, and 3.1 Å for $\gamma B7_{EC1-4}$ crystal form 2. Data collection and refinement statistics are given in Figure 1—source data 1.

Each of the Pcdh crystal structures consists of seven-strand beta sandwich EC domains arranged end to end, as expected, with three calcium ions bound at each of the EC–EC junctions by canonical cadherin family calcium-binding motifs. The structures are decorated with both N-linked glycans and O-linked mannoses (Figure 1A), including two EC2 G-strand O-linked mannoses (residues 193, 194, or 195 and 195, 196 or 197 in the various Pcdh structures), which appear to be conserved among clustered Pcdhs (Rubinstein et al., 2015; Goodman et al., 2016).

Flexibility in the γA *trans* dimer arrangement

The $\gamma B7_{EC1-4}$ structures each contain two molecules in the asymmetric unit, which are arranged in near identical anti-parallel EC1–4 mediated dimers (root mean square deviation over aligned C α atoms (RMSD) of 1.5 Å over 805 C α 's; Figure 1). These γB dimers are similar to those observed for the α - and β -Pcdh EC1–4 cell-cell recognition dimers we previously determined (Goodman et al., 2016) with pairwise RMSDs of 1.8–2.9 Å between α and γB dimer structures, and 3.3–4.0 Å between β and γB dimer structures (Figure 1—source data 3).

The γA structures showed an unanticipated variability in their molecular arrangement in the crystals. The $\gamma A1_{EC1-4}$ crystal structure contained four molecules in the asymmetric unit: Two of which are arranged in an EC1–4 mediated antiparallel dimer, with all four EC domains involved in the dimer interaction (chains A and B); and two are arranged in

an EC2–3 mediated antiparallel dimer, in which EC1 and EC4 are not involved in the dimer interaction (chains C and D). The EC2–3 portion of the dimer interaction is very similar between the two dimers in the structure (RMSD = 0.98 Å over 415 Cα's) and closely resembles the partial interaction observed in the previously published $\gamma A1_{EC1-3}$ structure (Nicoludis et al., 2015; Figure 1—figure supplement 1B). The main difference between the two dimers in the $\gamma A1$ crystal is therefore simply the presence or absence of the EC1:EC4 interaction. Since there are no protein domains filling the gap between EC1 and EC4 of chains C and D in the crystal, it is unclear why these domains do not interact. The fully engaged EC1–4-mediated dimer is similar to that of $\gamma B7_{EC1-4}$ and the published α - and β -Pcdh EC1–4-mediated dimers, involving the same interacting face of the molecule, however the RMSDs are quite large (4.3–5.0 Å; Figure 1—source data 3), highlighting the architectural differences between the $\gamma A1_{EC1-4}$ dimer and those of other Pcdh subtypes (Figure 1—figure supplement 1D and source data 2–4).

The $\gamma A8_{EC1-4}$ crystal structure contained a single molecule in the asymmetric unit, which is engaged with a symmetry mate in an anti-parallel EC2–3-mediated interaction involving the same surface of the molecule as in the other clustered Pcdh *trans* dimer structures. This crystal also contained a distinct interaction between symmetry-related molecules, also mediated by an anti-parallel EC2–3 interface and with a similar buried surface area (Figure 1—figure supplement 3). In order to confirm which interface is the biological *trans* dimerization interface, we generated a number of $\gamma A8$ arginine mutants separately targeting each of the observed interactions. Only those mutants that targeted the interaction surface in common with other Pcdhs resulted in loss of function in cell

aggregation assays (Figure 1B–C and Figure 1—figure supplement 3). It is this γ A8 dimer interaction that is shown in Figure 1A. Remarkably, like the γ A1_{EC1–4} dimer observed between chains C and D, the EC1:EC4 interaction is not formed. However, in the case of γ A8 the interaction surfaces of EC1 and EC4 instead interact with the EC4 domain of another symmetry-related molecule in the crystal.

Unexpectedly the γ A9_{EC1–5} crystal structure did not contain a *trans* dimer interaction in the crystal lattice. Given that γ A9_{EC1–5} is a low micromolar dimer in solution (Table 1), the monomeric arrangement in the crystal is likely an artifact of crystallization, perhaps due to the low pH (6.5) of the crystallization condition.

Both the γ A1_{EC1–4} and γ A8_{EC1–4} crystal structures contain dimers mediated solely by the EC2–3 regions of the *trans* interface, suggesting that for γ A Pcdhs the EC2–3 interaction might be sufficient for dimerization and cell-cell recognition. In addition, the crystal structure of γ A1_{EC1–3} (Nicoludis et al., 2015) contains the EC2–3 portion of the *trans* interaction. Whereas the published Pcdh EC1–3 structures from the β and C-type subfamilies (β 1, α C2, γ C3, and γ C5; Rubinstein et al., 2015; Nicoludis et al., 2015), did not contain any portion of the *trans* interface, and were monomeric in solution (Rubinstein et al., 2015). To determine whether the EC2–3 regions are sufficient for dimerization of γ -Pcdhs we produced EC1–3 fragments of two γ As and a γ B (γ A1, γ A4, and γ B6). However, AUC of these fragments showed that all three were monomeric in solution (Table 1), like the EC1–3 fragments of β 1 and γ C5 (Rubinstein et al., 2015).

Inter-family specificity

To understand why members of the α , β , γ A and γ B Pcdh subfamilies fail to form heterophilic complexes, we performed structural comparisons of the available homodimer structures. Excluding the γ A Pcdhs, which have diverse overall structures, the EC1–4 dimers of isoforms from the same subfamily have similar overall structures (RMSDs ~ 1.5 – 1.9 Å; Figure 1—source data 3). In contrast, superpositions of dimers from different subfamilies revealed much larger RMSDs due to distinct relative orientations of the individual domains (>3.7 Å; Figure 1—source data 3; Goodman et al., 2016; Nicoludis et al., 2016). This, in itself, provides a simple explanation for the absence of α/β , α/γ A, β/γ A, and β/γ B *trans* dimers. However, the four dimer structures from the α and γ B subfamilies exhibited intermediate structural similarity between the two subfamilies (RMSDs ~ 1.9 – 2.9 Å). We therefore sought to identify other conserved elements that might distinguish these subfamilies, and distinguish γ A and γ B Pcdhs, which are closely related in sequence.

The γ B7 structure reveals a salt bridge in the EC1:EC4 interface between residues E41 in EC1 and K338 in EC4 (Figure 2A). Both E41 and K338 are conserved in all γ B isoforms so that this salt bridge is likely present in all γ B homodimers (Figure 2C). In addition, residue R340, which is also conserved in all γ B isoforms, is positioned so that it could form an additional salt bridge with E41; however, no electron density was observed for its side chain. In contrast, the structures of γ A1 and γ A8 have an arginine or lysine at position 41 in the EC1 domain, which are conserved in all γ A isoforms (Figure 2C). Thus, a putative heterodimer formed between any γ B isoform and any isoform from γ A

would position a positively charged residue at position 41 in the EC1 domain of the γ A isoform in close proximity to K338 and R340 in the γ B isoform, which would significantly weaken binding (Figure 2A–C). Remarkably, α -Pcdhs also conserve a positive charge at position 40 (structurally equivalent to γ B E41), which suggests that putative heterodimers between α -Pcdhs and γ B-Pcdhs would also generate electrostatic clashes involving the same residues. Thus, the formation of heterodimers between γ B-Pcdhs and both γ A-Pcdhs and α -Pcdhs appears to be precluded by the conservation of key charged interface residues in EC1 and EC4. A similar mechanism was shown to determine family-wise specificity in the desmosomal cadherins (Harrison et al., 2016) and intra-family specificity in the case of nectins (Harrison et al., 2012).

Intra-family γ -Pcdh *trans*-recognition specificity

We next considered *trans*-recognition specificity among γ A isoforms and among γ B isoforms. Our previous analysis of α and β Pcdhs showed that interfacial residues that vary between isoforms, yet are conserved in orthologs of a given isoform, function as specificity determining residues (Goodman et al., 2016). Interactions between such residues were found to be favorable in homophilic complexes, but would typically generate steric or electrostatic clashes in potential heterophilic complexes. In order to identify specificity determining residues in γ B and γ A isoforms, we generated sequence logos derived from multiple sequence alignments of mammalian isoform-orthologs (Figure 3 and 4). The logo analysis reveals that the majority of isoform-specific *trans*-interface residues are highly conserved in the same isoform of other species.

To identify the likely roles of these residues in specificity we evaluated at the relationship between residues that interact across the *trans* interface. For example, in five γ B isoforms (γ B1– γ B5), contacting EC1:EC4 interface residues 75 and 367 are glutamate and arginine, which are conserved in orthologs and are likely to form a salt-bridge in the *trans* homodimer. However, in two isoforms (γ B6 and γ B7) these two residues are simultaneously changed to Q75 and N367. A hypothetical heterodimer between γ B7 and γ B3 would result in the unfavorable burial of two unsatisfied charged residues at the interface (Figure 3). Other examples of electrostatic compatibility/incompatibility for homophilic/heterophilic pairing can be seen in the interacting residues 111 and 298 of the EC2:EC3 interface of γ B isoforms (Figure 3), or the interacting residues 128 and 257 of the EC2:EC3 interface of γ A isoforms (Figure 4). We also found examples of small/large interacting residue pairs at the interface which showed correlated variations between isoforms such that heterophilic complexes would likely generate steric clashes. Such cases are found, for example, for residues 86 and 369 of the EC1:EC4 interface of γ B isoforms, residues 125 and 253 of the EC2:EC3 interface of γ B4 and γ B5 (Figure 3), and residues 79 and 340 of the EC1:EC4 interface of γ A8 and γ A9 (Figure 4). Finally, we identified the self-interacting residue 206 of γ A isoforms as a potential specificity-determining residue, providing hydrophobic contacts in some isoforms and polar contact in others (Figure 4).

***Cis* interactions**

EC6-dependent *cis* interactions of β - and γ B-Pcdhs, but not γ A-Pcdhs

We previously reported AUC data showing that γ B6, α C2 and γ C5 Pcdh EC1–6 fragments exist as dimers-of-dimers (tetramers) in solution, mediated by an EC1–4 interface and a distinct EC6-dependent interface (Rubinstein et al., 2015). Here, we have extended this analysis to determine the oligomeric states of multiple γ -Pcdh subfamily members and a representative of the β -Pcdh subfamily. All γ A EC1–6 molecules we tested formed dimers rather than tetramers in solution (Table 2). γ C3_{EC1–6} was also a dimer in solution, although in this case the isodesmic constant was only 1.5 fold larger than the dimer dissociation constant, indicating non-specific binding. These EC1–6 dimers are mediated by an EC1–4 (*trans*) interaction, since all the γ A and γ C3 EC1–4 fragments we measured were also dimers in solution (Table 1) and the γ A and γ C3 EC2–6 or EC3–6 fragments were monomers or very weak non-specific dimers (Table 2). In contrast, γ B6_{EC1–6}, γ C5_{EC1–6}, γ B2_{EC1–6} and β 5_{EC1–6} were tetrameric in solution (Table 2). In addition, the γ B2 EC3–6 fragment formed a dimer (Table 2), confirming the presence of the EC6-dependent *cis* interaction in solution for these γ B- and β -Pcdhs, in contrast to the γ A-Pcdhs. Since EC6 is highly conserved within non-C-type Pcdh subfamilies (average pairwise sequence identities for mouse EC6 domains are 90% for β s, 90% for γ As and 96% for γ Bs), we assume these results will be general to all mouse β , γ A and γ B isoforms.

We previously reported that γ A8_{EC2–6} was a dimer in solution (Rubinstein et al., 2015). However, it seems likely that this was due to the formation of an intermolecular

disulphide bond mediated by an exposed cysteine residue, as was observed in the γA_{EC1-3} crystal structure (Rubinstein et al., 2015).

Pcdh γ A carrier function suggests EC6-dependent heterophilic *cis* binding

We have shown that $\beta 17$, $\gamma B6$, $\alpha C2$, and $\gamma C5$ can interact with α -Pcdhs in an EC6-dependent manner (Thu et al., 2014). However, this has not been demonstrated for any γA isoform. Given the lack of a homophilic EC6-mediated homodimerization by γA isoforms in solution, we asked whether γA isoforms could interact heterophilically *in cis* with α -Pcdhs. To address this question, we performed the same assay as in Thu et al., 2014, which depends on the observation that α -Pcdhs are not delivered to the cell surface when expressed alone in K562 cells, and are therefore not able to mediate cell adhesion. α -Pcdhs require co-expression of an EC5–6-containing fragment of a ‘carrier’ Pcdh from another subfamily to reach the cell surface and mediate cell adhesion. We therefore tested whether non-adhesive EC5–6 containing fragments of $\gamma A3$ and $\gamma A9$ were able to deliver Pcdh $\alpha 4$ to the cell surface to mediate cell adhesion. Co-expression of both these isoform fragments with Pcdh $\alpha 4$ resulted in cell aggregation (Figure 5A) indicating that, despite their apparent lack of homophilic *cis* dimerization, γA -Pcdhs can interact heterophilically with α -Pcdhs *in cis*.

Crystal structures of γ -Pcdh EC3–6 fragments reveal the *cis*-interacting region

To further characterize γ -Pcdh *cis* interactions we sought to crystallize Pcdh fragments including both EC5, which may be involved in *cis* interactions, and the critical EC6 domain (Thu et al., 2014). From these experiments we obtained crystals of γA_{EC3-6} and γB_{EC3-6} , which diffracted to sufficient resolution for crystal structure determination. X-

ray diffraction by the $\gamma A4_{EC3-6}$ crystals was significantly anisotropic (Figure 5—figure supplement 1), and therefore anisotropic resolution limits were applied. The resolution of the final refined structures was 3.0/4.3/2.85 Å for $\gamma A4_{EC3-6}$ and 2.3 Å for $\gamma B2_{EC3-6}$. Data collection and refinement statistics are presented in Figure 5—source data 1.

Both the $\gamma A4_{EC3-6}$ and $\gamma B2_{EC3-6}$ crystal structures consisted of four EC domains connected by linkers, each containing three bound calcium ions as expected (Figure 5B). The two structures are similar overall (RMSD = 3.02 Å over 405 C α 's), although $\gamma A4_{EC3-6}$ shows a more pronounced EC4–EC5 bend angle (32.6° for $\gamma A4$ vs. 18.6° for $\gamma B2$). These are the first Pcdh structures containing EC6, which displays the classic beta sandwich fold, but with a large insertion between the A and A' strands (Figure 5C). This insertion is the one region of significant structural difference between the $\gamma A4$ and $\gamma B2$ EC6 domains, which otherwise have near identical structures (RMSD = 0.80 Å over 90 C α 's). Both structures are decorated with N- and O-linked sugar moieties throughout EC3–6, the majority of which are found on equivalent positions in both $\gamma A4$ and $\gamma B2$. Notably the G-strands of both EC6 domains are decorated with O-mannose groups on neighboring surface-facing residues, three for $\gamma B2$ and four for $\gamma A4$ (Figure 5B).

These EC3–6 structures, combined with the EC1–4 dimer structures, allowed us to model the EC1–6 *trans* dimer for γA and γB Pcdhs by structurally aligning the overlapping EC3–4 portions of the structures (Figure 5D–E). These models reveal an overall curved shape primarily defined by the EC4–5 bend angle, since both the EC1–4 dimer regions and the EC5–6 tails are relatively straight, and predict intermembrane spacing of ~360–375 Å.

The γA_{EC3-6} structure did not show any protein:protein interactions consistent with *cis* interactions in the crystal which, given that this γA family member is monomeric in solution (Table 2), was expected. However the γB_{EC3-6} crystal structure also did not reveal any interactions with clear biological relevance. Given that γB_{EC3-6} forms a weak *cis* dimer in solution (80.1 μM ; Table 2), this was unexpected. This monomeric arrangement in the crystal is likely an artifact of crystallization, perhaps due to the low pH of the crystallization condition (pH 6.5).

Mutagenesis experiments reveal the *cis*-interaction surface of γB EC6 domains

In order to identify the *cis* interface we carried out mutagenesis experiments using γB_6 , which has been shown to interact both homophilically and heterophilically *in cis* and behaves robustly in cell aggregation assays and in biophysical assays (Thu et al., 2014; Rubinstein et al., 2015; Table 2 and Figure 5A). We chose 11 EC6 surface residues, covering the entire surface of the domain, to mutate to aspartic acid. Wherever possible we chose residues that showed conserved differences between α -Pcdhs and other Pcdhs since it seemed likely that those residues account for the fact that α -Pcdhs do not form *cis* homodimers. We first tested the ability of these mutants to deliver an α -Pcdh to the cell surface. To accomplish this, we produced all eleven mutants in a non-adhesive $\Delta EC1$ γB_6 context. We confirmed these $\Delta EC1$ mutants to be non-adhesive in K562 cells when expressed alone, and then co-expressed each mutant with an α -Pcdh to determine whether the α -Pcdh was successfully delivered to the cell surface, as indicated by whether the α -Pcdh could mediate cell adhesion. The majority of the γB_6 mutants were able to deliver the α -Pcdh to the cell surface, but three mutants (L557D, V562D, and

R597D, γ B2 numbering) were not (Figure 6A). All three mutations mapped to the same surface of EC6, specifically to the B and E strands (Figure 6B)

We also assessed the behavior of these mutants in the full-length γ B6 context alone.

While most were still able to mediate cell aggregation like wild type γ B6 (Thu et al., 2014), the three mutants that were unable to deliver α -Pcdh to the cell surface in the Δ EC1 context were also unable to mediate cell aggregation in the full-length context (Figure 6A). Since all these mutations are in EC6 they should not affect the EC1–4-mediated *trans* interaction responsible for cell-cell adhesion in these assays. Thus, the fact that expression of these three mutants does not result in cell aggregation likely results from their failure to reach the cell surface.

To determine whether the L557D, V562D, and R597D γ B6 mutants disrupt the *cis* interface, we attempted to express them in the EC1–6 context to assess their oligomeric state in solution by AUC. We were only able to produce one of the mutants, V562D. This EC1–6 mutant was a dimer in solution rather than a (*cis-trans*) tetramer like the wild type (Figure 6C), indicating the V562D mutation did indeed disrupt homophilic *cis* interactions. These results also suggest that, like the α -Pcdhs (Thu et al., 2014), cell surface delivery of Pcdh γ B isoforms requires EC6-mediated *cis* interactions.

We used the PredUs2.0 program (Hwang et al., 2016), which combines structural homology with residue propensities to predict EC6 surface residues likely to participate in *cis* interactions (interface residues). Remarkably, all of the 23 residues predicted to be interfacial are located on one side of the molecule (Figure 6D and F) – the same side that was identified by mutagenesis. Together, these results allowed us to define a putative *cis*

interaction region that encompasses the A, B, D, and E strands and the BC and DE loops (Figure 6E and F). Sequence alignment of the EC6 domains for α , β , and γ isoforms shows that α -Pcdhs and the carrier β - and γ -Pcdhs differ in nine residues in this region (Figure 6F). The nine residues group into three clusters in the putative *cis* interface. Right side cluster: residues 530, 534 (A strand), and 562 (B strand), middle cluster: residues 556, 556 (B strand), 588 (D strand), and 597 (E strand), and the top cluster: residues 570 (BC loop) and 592 (DE loop) (Figure 6E). The three *cis*-disruptive mutants (L557D, V562D, and R597D) are mapped onto two of these clusters indicating that they participate in Pcdh *cis* interactions. In contrast, O-mannosylation is observed in the structures at residues 624, 626, and 628 (γ B2 numbering) in the EC6 domain G-strands of both the γ B2 and γ A4 structures – on the opposite molecular face to the mutations that disrupt cell surface delivery (Figure 6B). These positions are usually conserved in α , β , and γ Pcdhs (Figure 6F) suggesting that these O-glycans are likely present in all alternate Pcdhs.

The structural basis for the differences in homophilic *cis* binding observed for γ A and γ B/ β isoforms is not as clear. However, conserved sequence differences in the DE loop region between the γ A, γ B and β subfamilies— γ A = GLHT, γ B = GLRT, and β = WAHN—as well as the top of the A strand (adjacent to B strand residue 562)—residues 531–532 = EI in γ A, RV in γ B, and FV in β isoforms —could contribute to the different subfamily *cis* interaction characteristics (Figure 6—figure supplement 1).

Discussion

The structures of representative γ A and γ B Pcdh protein isoforms reported here complete a set of representative structures for *trans*-recognition interfaces from alternate clustered Pcdh isoforms, with structures now available for at least two Pcdhs from each of the α , β (Goodman et al., 2016), γ B (Nicoludis et al., 2016; this paper) and γ A (this paper) subfamilies. Representative structures of engaged *trans* dimers of C-type Pcdhs have yet to be obtained. As discussed below, the collection of protocadherin structures now available present a clear picture of how *trans*-homodimeric interaction specificity is coded for alternate Pcdh isoforms on the *trans* dimer interface comprising domains EC1–EC4. We also report a monomeric structure of a region containing the *cis*-interacting EC6 domain, and use it, together with mutagenesis experiments, to locate the *cis* interface in Pcdhs. In addition, our data indicate that γ A and γ B isoforms are distinct subfamilies with regard to their *cis* and *trans* protein interactions. With this information in hand, we discuss alternate mechanisms that have been proposed for the molecular basis of Pcdh-mediated neuronal self-recognition and non-self discrimination.

Pcdh *trans* interaction specificity

The homophilic recognition properties of alternate (non C-type) clustered Pcdhs may be understood at the subfamily and isoform levels. Members of different subfamilies fail to bind to each other in *trans* primarily due to structural differences between the α , β and γ A subfamilies. That is, the putative dimers they would form would not exhibit shape compatibility. However, members of the γ B subfamily are sufficiently similar in structure to members of the α subfamily that a specificity mechanism is unlikely to be based entirely on shape complementarity. However, the sequence and structural analyses presented above show that that EC1:EC4 interface in γ B isoforms will contain salt

bridges in the homodimers, whereas the comparable interaction in the inter-subfamily heterodimer would lead to incompatible electrostatic repulsion. In addition, electrostatic clashes involving the same residues appear to preclude formation of heterodimers between γ B-Pcdhs and γ A-Pcdhs. These then are cases where subfamily level specificity is encoded in the EC1:EC4 interface.

Sequence and structural analyses also identify the determinants of intra-subfamily specificity. In agreement with our previous analysis of the α and β Pcdhs (Goodman et al., 2016) we find that the electrostatic and steric compatibility apparent in homodimer structures would be replaced by incompatibility in putative heterodimers. As discussed above, some of the specificity-determining interactions are located in the EC1:EC4 interface and some in the EC2:EC3 interface. These findings, as well as those summarized in the previous paragraph contradict a primary conclusion reached by Nicoludis et al. (2016). Based on their structure of the EC1-4 *trans* dimer of Pcdh γ B3 and of the four *trans*-dimeric α and β isoform structures we previously determined (Nicoludis et al., 2016; Goodman et al. 2016), they used a bioinformatics analysis to infer that *trans* interaction specificity is mediated by the EC2:EC3 interaction, and that the EC1:EC4 interaction provides affinity, but not specificity. Our analysis, in contrast, reveals numerous specificity elements in EC1:EC4 interactions.

The importance of the EC1:EC4 interaction to *trans*-binding specificity is also demonstrated by our published experimental results with Pcdh mutants. Cell aggregation experiments with domain-shuffled mutants have clearly demonstrated that specificity is dependent on the identity of EC1 and EC4 (Supplementary Figure 3 in Rubinstein et al.,

2015). In the case of α -Pcdhs, it is true that the EC1 and EC4 interface residues are mainly conserved between isoforms as we previously reported (Goodman et al, 2016). However, some isoforms show conserved differences that determine specificity: $\alpha 7$ shows isoform-specific conservation of *trans*-interface EC1 residues 36 and 38 and EC4 residues 322 and 324. Most importantly, swapping these residues between $\alpha 7$ and $\alpha 8$ swaps their recognition specificities (Rubinstein et al, 2015). For β , γA , and γB Pcdhs, isoform-specific conservation of EC1:EC4 *trans*-interface residues is observed in almost all isoforms, as can be seen in sequence logo analysis (Figure 3, Figure 4, and Goodman et al, 2016 Figure 2). This observation, in addition to results from functional mutagenesis experiments which show changes in specificity when such residues are mutated (Goodman et al, 2016), clearly demonstrate that both the EC2:EC3 and EC1:EC4 interfaces play important roles in determining binding specificity.

EC6 domain structure and *cis* interactions

Pcdh *cis* multimers have been suggested to form promiscuously between isoforms, and to thereby diversify the functional Pcdh repertoire (Schreiner & Weiner 2010; Yagi et al., 2013; Thu et al, 2014; Rubinstein et al, 2015). We previously used domain-deletion studies of numerous Pcdh isoforms to localize the *cis* interaction region to the EC6 domain, with possible contributions from EC5, and showed that the *cis* complexes formed are dimeric (Rubinstein et al., 2015). Here we report structures containing monomeric EC6 domains, and locate their dimeric recognition regions by identifying mutations that interfere with the formation of *cis* dimers for both α - and γB -Pcdhs (Figure 6).

Sequence comparisons of Pcdh EC6 domains (Figure 6 and Figure 6—figure supplement 1) revealed conserved differences between the Pcdh subfamilies, which are likely to relate to their *cis*-interaction specificities (Thu et al, 2014, Rubinstein et al, 2015). We previously showed subfamily specific diversity in *cis* interactions in that α -Pcdhs and Pcdh γ C4 are not transported alone to the cell surface, but only when engaged in *cis*-dimeric complexes with ‘carrier’ Pcdhs corresponding to other isoforms, including alternate β and γ , and some C-type Pcdhs (Thu et al, 2014). Our results suggest additional diversity in Pcdh *cis* interactions: we found through biophysical measurements that two alternate γ B-Pcdhs interacted homophilically *in cis* in solution (Table 2), but three alternate γ A Pcdhs did not. In light of the high level of sequence conservation of the EC6 domains within the γ A and within the γ B Pcdh subfamilies, it is likely that, in general, γ A Pcdhs fail to dimerize or form only weak *cis* dimers (enabled in part by the constrained 2D environment of the membrane surface (Wu et al., 2011)), while alternate γ B/ γ B *cis*-dimers are expected to have significant affinity. Despite the difference in dimerization affinities, both the γ A and γ B Pcdhs functioned as carriers for α -Pcdhs (Figure 5A), and show high sequence conservation (Figure 6—figure supplement 1) consistent with the participation of both γ A and γ B isoforms in Pcdh *cis* dimers. Overall, these observations clearly show an unanticipated specificity in *cis*-dimer formation.

Implications for neuronal recognition

Subfamily specific differences in *cis*-dimerization specificity are expected to impact the diversity and composition of the functional Pcdh repertoire of *cis*-dimeric recognition units. Instead of being composed of random isoform combinations as previously

suggested (Yagi et al., 2013; Thu et al., 2014; Rubinstein et al., 2015), the repertoire composition of *cis*-dimers is predicted to be non-uniform assuming comparable levels of monomers. For example, no recognition units consisting of two alternative α -Pcdh isoforms are expected to form, and $\gamma A/\gamma A$ recognition units would be absent or less frequent, than $\gamma B/\gamma B$ recognition units. Since the composition of the *cis*-dimeric repertoire is limited compared to all random combinations, the recognition-unit diversity encoded by stochastic expression of Pcdh isoforms is likely to be less than previously thought.

We have previously described two alternative molecular mechanisms for neuronal self-recognition through *trans* interactions of Pcdh *cis*-dimeric recognition units (Thu et al., 2014; Rubinstein et al., 2015). Both of these mechanisms depend on diverse repertoires of dimeric recognition units to achieve sufficient levels of cell surface diversity such that non-self neurons are not inappropriately recognized as self. In the first case (Figure 7B), *trans* binding is envisioned to occur only between recognition units with precisely matched isoform composition, and results in the formation of a dimer-of-dimers containing maximally two Pcdh isoforms. As we described previously (Thu et al, 2014), this model leads to a limited number of possible distinct cell surface identities and even fewer if the population of *cis* dimers is not random. In the second case (Figure 7C), *trans* binding is suggested to occur between recognition units with a single matched isoform, resulting in the formation of a zipper or chain of Pcdh dimers arrayed between membrane surfaces (Thu et al, 2014); the chain of dimeric recognition units is proposed to be terminated by the presence of a single mismatched isoform (Figure 7C). This chain termination model leads to the ability to encode a far larger set of distinct cell surface

identities (Rubinstein et al., 2015). However, to date there has been no direct observation of oligomeric Pcdhs on cell surfaces. Since it remains unclear what proportion of neurons utilize Pcdhs for self-avoidance, and thus the Pcdh diversity required to avoid inappropriate self-recognition of interacting neurons remains unclear; we cannot currently distinguish between these models.

Materials and Methods

Protein production

cDNAs for Pcdh ectodomain fragments, excluding the predicted signal sequences, were cloned into a p α SHP-H mammalian expression vector (a kind gift from Daniel J. Leahy, John Hopkins University) modified with the BiP signal sequence and a C-terminal octahistidine tag (Rubinstein et al., 2015). The signal sequences were predicted using the SignalP 4.0 server (Petersen et al., 2011).

Suspension-adapted HEK293 Freestyle cells (Invitrogen) in serum free media (Invitrogen) were used for protein expression. The plasmid constructs were transfected into cells using polyethyleneimine (Polysciences Inc.). Media was harvested ~6 days after transfection and the secreted proteins were purified by nickel affinity chromatography followed by size exclusion chromatography in 10 mM Tris pH 8.0, 150 mM sodium chloride, 3 mM calcium chloride, and 100–250 mM imidazole pH 8.0. Purified proteins were concentrated to >2 mg/ml and used for analytical ultracentrifugation or crystallization experiments.

Sedimentation equilibrium analytical ultracentrifugation (AUC)

Experiments were performed in a Beckman XL-A/I analytical ultracentrifuge (Beckman-Coulter, Palo Alto CA, USA), utilizing six-cell centerpieces with straight walls, 12 mm path length and sapphire windows. Samples were dialyzed overnight and then diluted 10 mM Tris pH 8.0, 150 mM NaCl, 3 mM CaCl₂ with varying concentration of imidazole pH 8.0, as follows: 100 mM (α 7_{EC1-5}/ γ C3_{EC6} chimera), 200 mM (β 5_{EC1-6}, γ B2_{EC1-6}, γ B2_{EC3-6}, γ A1_{EC3-6}, γ A4_{EC1-4}, γ B5_{EC1-4}, γ B7_{EC1-4}, γ C3_{EC1-4}), and 250 mM (γ A1_{EC1-4},

$\gamma A1_{EC2-6}$, $\gamma A1_{EC1-6}$, $\gamma A4_{EC1-3}$, $\gamma A4_{EC3-6}$, $\gamma A4_{EC1-6}$, $\gamma A9_{EC1-5}$, $\gamma A9_{EC1-6}$, $\gamma B6_{EC1-3}$, $\gamma B6_{EC1-6}$, $\gamma C3_{EC3-6}$, $\gamma C3_{EC1-6}$). Proteins were diluted to an absorbance at 10 mm and 280 nm of 0.65, 0.43 and 0.23 in channels A, B and C, respectively. The dilution buffer was used as blank. All samples were run at four speeds, the lowest speed was held for 20 h then four scans with 1 h interval, the subsequent three speeds were each held for 10 h followed by four scans with 1h interval. The speeds were 9000, 11000, 13000 and 15000 rpm (all $EC1-6$, $EC2-6$ and $EC1-5$ constructs) or 11000, 14000, 17000 and 20000 rpm (all $EC1-3$, $EC1-4$ and $EC3-6$ constructs). Measurements were done at 25°C, and detection was by UV at 280 nm. Solvent density and protein v -bar at both temperatures were determined using the program SednTerp (Alliance Protein Laboratories, Corte Cancion, Thousand Oaks, CA, USA). For calculation of dimeric K_d and apparent molecular weight, all useful data were used in a global fit, using the program HeteroAnalysis, obtained from University of Connecticut. (www.biotech.uconn.edu/auf). Calculation of the tetramer K_d s was done with the program Sedphat (<http://www.analyticalultracentrifugation.com/sedphat/index.htm>).

Crystallization and X-ray data collection

Protein crystals were grown using the vapor diffusion method. Crystallization conditions were as follows, with cryo-protectants used given in parentheses: 8% (w/v) PEG8000, 16% ethylene glycol, 20% Morpheus Amino Acids (Molecular Dimensions), 0.1 M Morpheus Buffer System 2 (Hepes/MOPS buffer; Molecular Dimensions) pH 7.0 for $\gamma A1_{EC1-4}$; 11% isopropanol, 50 mM sodium chloride, 0.1 M Hepes pH 7.5 (30% ethylene glycol) for $\gamma A8_{EC1-4}$; 10% (w/v) PEG4000, 20% (v/v) glycerol, 30 mM magnesium chloride, 30 mM calcium chloride, 0.1 M Morpheus Buffer System 1 (Mes/Imidazole

buffer; Molecular Dimensions) pH 6.5 for $\gamma A9_{EC1-5}$; 0.1 M Tris-Cl pH 8.5, 0.2 M trimethylamine N-oxide, 3% dextran sulfate sodium salt 5000, 17% (w/v) PEG2000MME (20% (v/v) glycerol) for $\gamma B7_{EC1-4}$ crystal form 1; 0.1 M Tris-Cl pH 8.5, 0.2 trimethylamine N-oxide, 5% (v/v) Jeffamine M-600 pH 7.0, 17% (w/v) PEG2000MME (20% (v/v) PEG400) for $\gamma B7_{EC1-4}$ crystal form 2; 0.1 M ammonium sulfate, 9% (w/v) PEG20000, 18% PEG550MME, 0.1 M Morpheus Buffer System 3 (Tris/Bicine; Molecular Dimensions) pH 8.5 for $\gamma A4_{EC3-6}$; 11.5% (w/v) PEG8000, 23% (v/v) ethylene glycol, 30 mM magnesium chloride, 30 mM calcium chloride, 0.1 M Morpheus Buffer System 1 (Mes/Imidazole buffer; Molecular Dimensions) pH 6.5 for $\gamma B2_{EC3-6}$. X-ray diffraction data was collected at 100K from single crystals at Northeastern Collaborative Access Team (NE-CAT) beamlines 24ID-C and 24ID-E at the Advanced Photon Source, Argonne National Laboratory. All datasets were indexed using XDS (Kabsch, 2010) and initially scaled using AIMLESS (Evans, 2006; Evans and Murshudov, 2013), except the $\gamma A8_{EC1-4}$ data which was indexed with iMOSFLM (Battye et al., 2011) and scaled using SCALA (Evans, 2006).

Diffraction anisotropy and pseudosymmetry

The $\gamma A9_{EC1-5}$, $\gamma B7_{EC1-4}$ crystal form 1, and $\gamma A4_{EC3-6}$ diffraction data all showed strong diffraction anisotropy, with much weaker diffraction along a^* or b^* or both (Figure 1—figure supplement 2 and Figure 5—figure supplement 1). These data were therefore truncated using ellipsoidal limits with using a 3.0 F/sigma cut-off along each of the three principle crystal axes as implemented in the UCLA Diffraction Anisotropy Server (Strong et al., 2006). However we did not use the server's default scaling procedure to remove anisotropy from the data in the final rounds of refinement. Instead an overall

anisotropic B-factor was applied to the model by Phenix (Adams et al., 2010), as is standard, during refinement to account for the data anisotropy.

The $\gamma B2_{EC3-6}$ diffraction data showed translational pseudosymmetry with a large Patterson peak (60.9% height relative to the origin) at 0.000, 0.000, 0.323. This likely affected the intensity statistics and it is possible this also led to the higher R-values obtained in refinement: Final R_{work}/R_{free} (24.78%/27.78%) were higher than is common for a 2.3 Å dataset despite the apparent high quality of the electron density map.

Crystal structure phasing and refinement

All structures were solved by molecular replacement using Phaser (McCoy et al., 2007): $\gamma A1_{EC1-4}$ was solved using the $\gamma A1_{EC1-3}$ structure (PDB: 4ZI9) as a search model; $\gamma A8_{EC1-4}$ was solved using $\gamma A8_{EC1-3}$ (PDB: 4ZPS); $\gamma A9_{EC1-5}$ was solved using EC2–3 of $\gamma A8_{EC1-4}$; $\gamma B7_{EC1-4}$ was solved using ensembles of individual Pcdh EC domains from multiple isoform structures; $\gamma B2_{EC3-6}$ was solved using EC3–5 from the $\alpha 7_{EC1-5}$ structure (PDB: 5DZV); and $\gamma A4_{EC3-6}$ was solved using EC3–4 from $\gamma A8_{EC1-4}$, EC5 from $\gamma A9_{EC1-5}$ and EC6 from $\gamma B2_{EC3-6}$.

Iterative model building using Coot (Emsley et al., 2010) and maximum-likelihood refinement using Phenix (Adams et al., 2010) was conducted yielding the final refined structures whose statistics are reported in Figure 1—source data 1 and Figure 5—source data 1.

The electron density maps obtained were generally of reasonable quality, however the $\gamma B7_{EC1-4}$ crystal form 2 map had poor density for the bottom half of EC4 in chain B and

the neighboring top half of EC1 in chain A. Side chains were not observed in the map for many of the residues in these regions and were therefore not built. The density for EC4 in chain A and EC1 in chain B, including the interfacial regions was much better. The $\gamma A9_{EC1-5}$ map showed poor electron density for EC1, and the $\gamma A4_{EC3-6}$ map showed poor density for EC3. In addition the $\gamma A1_{EC1-4}$, $\gamma A8_{EC1-4}$, and $\gamma B7_{EC1-4}$ crystal form 1 structures were all very low resolution, at 4.2 Å, 3.6 Å, 4.5/4.5/3.6 Å respectively, and therefore many of the side chain positions/rotamers were not clearly defined in the electron density map. We therefore limited our analysis of the interfacial regions of these molecules to looking at which residues were in close proximity rather than the precise atomic arrangements.

Structure analysis

UCSF Chimera (Pettersen et al., 2004) was used to generate unmodeled side chains using the Dunbrack rotamer library prior to buried surface area (BSA) calculations. BSAs are given as the change in accessible surface area over both protomers and were calculated using 'Protein interfaces, surfaces and assemblies' service (PISA) at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html; Krissinel and Henrick, 2007). Interdomain angles were calculated using UCSF Chimera. Root mean square deviations over aligned C α atoms between structures were calculated using Pymol (Schrödinger, LLC). Crystal structure figures were made using Pymol.

Generation of Pcdh isoform sequence conservation logos

Orthologs of the mouse γA and γB Pcdh isoforms were collected from an annotation pipeline link at the NCBI database (Wheeler et al., 2008). Blast (Altschul et al., 1997)

was used to filter out any candidate orthologs with significant similarity to more than one mouse Pcdh isoform. The species for which we identified orthologs of the mouse γ A and γ B Pcdh isoforms are listed in Figure 3—source data 1 and Figure 4—source data 1. Multiple sequence alignments were generated using Clustal Omega (Sievers et al., 2011) and sequence logos were generated using WebLogo3 (Crooks et al., 2004).

Cell aggregation assay to test *trans* binding mutants

A pMax expression construct encoding full-length Pcdh- γ A8 with a C-terminal mCherry-tag was used as described in Thu et al., 2014. Mutants were generated using the Quikchange method (Stratagene). Cell aggregation assays were performed as previously described in Thu et al., 2014. Briefly, the Pcdh expression constructs were transfected into K562 cells (human leukemia cell line, ATCC CCL243) by electroporation using an Amaxa 4D-Nucleofactor (Lonza). After 24 hours, the transfected cells were mixed by shaking for one to three hours. The cells were then imaged with an Olympus fluorescent microscope to determine whether or not they had aggregated.

Co-transfection assays testing cell surface delivery of α -Pcdhs by other Pcdhs and mutants

Co-transfection assays were performed as previously described in Thu et al., 2014 and in a similar manner to the cell aggregation assays described above. C-terminal mCherry-tagged constructs of full length Pcdh α 4 or Pcdh α 7 were co-transfected with C-terminal mCherry-tagged constructs of various Δ EC1 Pcdhs and Pcdh mutants into K562 cells by electroporation as described above. Transfected cells were mixed by shaking for 1–3 hours and then imaged to see whether they had aggregated, as described above. Each

construct was also transfected into K562 cells alone to confirm that both the Δ EC1 Pcdhs and the α -Pcdhs could not mediate cell aggregation when expressed alone, as previously observed (Thu et al., 2014).

Accession numbers

Atomic coordinates and structure factors are deposited in the protein data bank with accession codes PDB: 5SZL, 5SZM, 5SZN, 5SZO, 5SZP, 5SZQ, and 5SZR.

Author contributions

K.M.G., R.R., T.M., B.H., and L.S. designed experiments, analyzed data and wrote the paper. S.M., F.B., and K.M.G. cloned, expressed, purified and crystallized the proteins. K.M.G. determined the crystal structures. S.M. performed the site-directed mutagenesis. R.R. conducted the sequence and structural analysis. C.A.T. and C.R. performed and analyzed the cell aggregation experiments. G.A. performed and analyzed the analytical ultracentrifugation experiments.

Acknowledgements

We thank Surajit Banerjee, Igor Kourinov, David Neau, and Frank V. Murphy for help with synchrotron data collection at the APS NE-CAT 24-ID-C/E beamlines, supported by NIH P41 GM103403. This work was supported by a National Science Foundation grant to B.H. (MCB- 1412472), an NIH grant to L.S. (R01GM062270), and a joint NIH grant to T.M. and L.S. (R01GM107571).

Competing Interests

The authors declare no competing interests.

Supplementary Files

Figure 1—figure supplements 1–3 and source data 1–5

Figure 2—figure supplement 1

Figure 3—figure supplement 1 and source data 1

Figure 4—figure supplement 1 and source data 1

Figure 5—figure supplement 1 and source data 1

Figure 6—figure supplement 1

References

- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., and Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D* 66, 213–221. doi: 10.1107/S0907444909052925
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst. D* 67, 271–281. doi: 10.1107/S0907444910048675
- Chen, W.V., Alvarez, F.J., Lefebvre, J.L., Friedman, B., Nwakeze, C., Geiman, E., Smith, C., Thu, C.A., Tapia, J.C., Tasic, B., Sanes, J.R., and Maniatis, T. (2012). Functional significance of isoform diversification in the protocadherin gamma gene cluster. *Neuron* 75, 402–9. doi: 10.1016/j.neuron.2012.06.039
- Chen, W.V., and Maniatis, T. (2013). Clustered protocadherins. *Development* 140, 3297–3302. doi: 10.1242/dev.090621
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner S.E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Cryst. D* 66, 486–501. doi: 10.1101/gr.849004
- Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., Koide, T., Kitsukawa, T., Hamada, S., and Yagi, T. (2005). Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. *Nat. Genet.* 37, 171–176. doi:10.1038/ng1500
- Evans P.R. (2006). Scaling and assessment of data quality. *Acta Cryst. D* 62, 72–82. doi: 10.1107/S0907444905036693
- Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Cryst. D* 69, 1204–1214. doi: 10.1107/S0907444913000061
- Garrett, A.M., Schreiner, D., Lobas, M.A., and Weiner, J.A. (2012). γ -protocadherins control cortical dendrite arborization by regulating the activity of a FAK/PKC/MARCKS signaling pathway. *Neuron* 74, 269–76. doi: 10.1016/j.neuron.2012.01.028
- Goodman, K.M., Rubinstein, R., Thu, C.A., Bahna, F., Mannepalli, S., Ahlsén, G., Rittenhouse, C., Maniatis, T., Honig, B., and Shapiro, L. (2016). Structural Basis of Diverse Homophilic Recognition by Clustered α - and β -Protocadherins. *Neuron* 90, 709–723. doi:10.1016/j.neuron.2016.04.004

- Harrison, O.J., Vendome, J., Brasch, J., Jin, X., Hong, S., Katsamba, P.S., Ahlsen, G., Troyanovsky, R.B., Troyanovsky, S.M., Honig, B., and Shapiro, L. (2012). Nectin ectodomain structures reveal a canonical adhesive interface. *Nat. Struct. Mol. Biol.* *19*, 906–15. doi: 10.1038/nsmb.2366
- Harrison, O.J., Brasch, J., Lasso, G., Katsamba, P.S., Ahlsen, G., Honig, B., and Shapiro, L. (2016). Structural basis of adhesive binding by desmocollins and desmogleins. *Proc. Natl. Acad. Sci. USA* *113*, 7160–5. doi: 10.1073/pnas.1606272113
- Hasegawa, S., Hamada, S., Kumode, Y., Esumi, S., Katori, S., Fukuda, E., Uchiyama, Y., Hirabayashi, T., Mombaerts, P., and Yagi, T. (2008). The protocadherin-alpha family is involved in axonal coalescence of olfactory sensory neurons into glomeruli of the olfactory bulb in mouse. *Mol. Cell Neurosci.* *38*, 66–79. doi: 10.1016/j.mcn.2008.01.016
- Hasegawa, S., Hirabayashi, T., Kondo, T., Inoue, K., Esumi, S., Okayama, A., Hamada, S., and Yagi, T. (2012). Constitutively expressed Protocadherin- α regulates the coalescence and elimination of homotypic olfactory axons through its cytoplasmic region. *Front Mol. Neurosci.* *5*, 97. doi: 10.3389/fnmol.2012.00097
- Hirano, K., Kaneko, R., Izawa, T., Kawaguchi, M., Kitsukawa, T., and Yagi, T. (2012). Single-neuron diversity generated by Protocadherin-beta cluster in mouse central and peripheral nervous systems. *Front Mol. Neurosci.* *5*, 90. doi: 10.3389/fnmol.2012.00090
- Hwang, H., Petrey, D., and Honig, B. (2016). A hybrid method for protein-protein interface prediction. *Protein Sci.* *25*, 159–65. doi: 10.1002/pro.2744
- Kabsch W. (2010) XDS. *Acta Cryst. D* *66*, 125–132. doi: 10.1107/S0907444909047337
- Kaneko, R., Kato, H., Kawamura, Y., Esumi, S., Hirayama, T., Hirabayashi, T., and Yagi, T. (2006). Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. *J Biol. Chem.* *281*, 30551–3560. doi: 10.1074/jbc.M605677200
- Katori, S., Hamada, S., Noguchi, Y., Fukuda, E., Yamamoto, T., Yamamoto, H., Hasegawa, S., and Yagi, T. (2009). Protocadherin-alpha family is required for serotonergic projections to appropriately innervate target brain areas. *J. Neurosci.* *29*, 9137–47. doi: 10.1523/JNEUROSCI.5478-08.2009
- Kostadinov, D., and Sanes, J.R. (2015). Protocadherin-dependent dendritic self-avoidance regulates neural connectivity and circuit function. *Elife* *4*, e08964. doi: 10.7554/eLife.08964
- Krissinel, E. and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* *372*, 774–797. doi:10.1016/j.jmb.2007.05.022
- Lefebvre, J.L., Zhang, Y., Meister, M., Wang, X., and Sanes, J.R. (2008). Gamma-protocadherins regulate neuronal survival but are dispensable for circuit formation in retina. *Development.* *135*, 4141–51. doi: 10.1242/dev.027912
- Lefebvre, J.L., Kostadinov, D., Chen, W.V., Maniatis, T., and Sanes, J.R. (2012). Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* *488*, 517–521. doi: 10.1038/nature11305

- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *Journal of applied crystallography* *40*, 658–674. doi: 10.1107/S0021889807021206
- Molumby, M.J., Keeler, A.B., and Weiner, J.A. (2016). Homophilic Protocadherin Cell-Cell Interactions Promote Dendrite Complexity. *Cell Reports* *15*, 1037–50. doi: 10.1016/j.celrep.2016.03.093.
- Miura, S.K., Martins, A., Zhang, K.X., Graveley, B.R., and Zipursky, S.L. (2013). Probabilistic splicing of *Dscam1* establishes identity at the level of single neurons. *Cell* *155*, 1166–1177. doi: 10.1016/j.cell.2013.10.018
- Neves, G., Zucker, J., Daly, M., and Chess, A. (2004). Stochastic yet biased expression of multiple *Dscam* splice variants by individual cells. *Nat. Genet.* *36*, 240–246.
- Nicoludis, J.M., Lau, S.Y., Schärfe, C.P.I., Marks, D.S., Weihofen, W.A., and Gaudet, R. (2015). Structure and Sequence Analyses of Clustered Protocadherins Reveal Antiparallel Interactions that Mediate Homophilic Specificity. *Structure*. *23*, 2087–98. doi: 10.1016/j.str.2015.09.005
- Nicoludis, J.M., Vogt, B.E., Green, A.G., Schärfe, C.P., Marks, D.S., and Gaudet, R. (2016). Antiparallel protocadherin homodimers use distinct affinity-and specificity mediating regions in cadherin repeats 1-4. *eLife* doi: 10.7554/eLife.18449
- Petersen, T.N., Brunak S., von Heijne G., Nielsen H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* *8*(10), 785–786. doi: 10.1038/nmeth.1701
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* *25*, 1605–1612. doi: 10.1002/jcc.20084
- Prasad, T., and Weiner, J.A. (2011). Direct and Indirect Regulation of Spinal Cord Ia Afferent Terminal Formation by the γ -Protocadherins. *Front Mol. Neurosci.* *4*, 54. doi: 10.3389/fnmol.2011.00054
- Rubinstein, R., Thu, C.A., Goodman, K.M., Wolcott, H.N., Bahna, F., Manneppalli, S., Ahlsén, G., Chevee, M., Halim, A., Clausen, H., Maniatis, T., Shapiro, L., and Honig, B. (2015). Molecular logic of neuronal self-recognition through protocadherin domain interactions. *Cell* *163*, 629–642. doi: 10.1016/j.cell.2015.09.026
- Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E., and Zipursky, S.L. (2000). *Drosophila Dscam* is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* *101*, 671–684. doi:10.1016/S0092-8674(00)80878-8
- Schreiner, D., and Weiner, J.A. (2010). Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc. Natl. Acad. Sci. USA* *107*, 14893–14898. doi: 10.1073/pnas.1004526107
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., and Higgins, D.G. (2011).

- Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 7, 539. doi: 10.1038/msb.2011.75
- Strong M., Sawaya M.R., Wang S., Phillips M., Cascio D., Eisenberg D. (2006). Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* 103, 8060–8065. doi: 10.1073/pnas.0602606103
- Suo, L., Lu, H., Ying, G., Capecchi, M.R., and Wu, Q. (2012). Protocadherin clusters and cell adhesion kinase regulate dendrite complexity through Rho GTPase. *J. Mol. Cell Biol.* 4, 362–76. doi: 10.1093/jmcb/mjs034
- Tasic, B., Nabholz, C.E., Baldwin, K.K., Kim, Y., Rueckert, E.H., Ribich, S.A., Cramer, P., Wu, Q., Axel, R., and Maniatis, T. (2002). Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Mol. Cell* 10, 21–33. doi: 10.1016/S1097-2765(02)00578-6
- Thu, C.A., Chen, W.V., Rubinstein, R., Chevee, M., Wolcott, H.N., Felsovalyi, K.O., Tapia, J.C., Shapiro, L., Honig, B., and Maniatis, T. (2014). Single-cell identity generated by combinatorial homophilic interactions between alpha, beta, and gamma protocadherins. *Cell* 158, 1045–1059. doi: 10.1016/j.cell.2014.07.012
- Wang, X., Su, H., and Bradley, A., (2002). Molecular mechanisms governing Pcdh-gamma gene expression: evidence for a multiple promoter and cis-alternative splicing model. *Genes Dev.* 16, 1890–1905. doi: 10.1101/gad.1004802
- Weiner, J.A., Wang, X., Tapia, J.C., and Sanes, J.R. (2005). Gamma protocadherins are required for synaptic development in the spinal cord. *Proc. Natl. Acad. Sci. USA* 102, 8–14. doi: 10.1073/pnas.0407931101
- Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Edgar, R., Federhen, S., *et al.* (2008). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 36, D13–21. doi: 10.1093/nar/gkv1290
- Wojtowicz, W.M., Flanagan, J.J., Millard, S.S., Zipursky, S.L., and Clemens, J.C. (2004). Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* 118, 619–633. doi: 10.1016/j.cell.2004.08.021
- Wojtowicz, W.M., Wu, W., Andre, I., Qian, B., Baker, D., and Zipursky, S.L. (2007). A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. *Cell* 130, 1134–1145. doi:10.1016/j.cell.2007.08.026
- Wu, Q., and Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 97, 779–790. doi:10.1016/S0092-8674(00)80789-8
- Wu, Q., Zhang, T., Cheng, J.F., Kim, Y., Grimwood, J., Schmutz, J., Dickson, M., Noonan, J.P., Zhang, M.Q., Myers, R.M., and Maniatis, T. (2001). Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Res.* 11, 389–404. doi: 10.1101/gr.167301

Wu, Y., Vendome, J., Shapiro, L., Ben-Shaul, A., and Honig, B. (2011). Transforming binding affinities from three dimensions to two with application to cadherin clustering. *Nature*. 475, 510–3. doi: 10.1038/nature10183

Yagi, T. (2013). Genetic basis of neuronal individuality in the mammalian brain. *J. Neurogenet.* 27, 97–105. doi: 10.3109/01677063.2013.801969

Zhan, X.L., Clemens, J.C., Neves, G., Hattori, D., Flanagan, J.J., Hummel, T., Vasconcelos, M.L., Chess, A., and Zipursky, S.L. (2004). Analysis of Dscam diversity in regulating axon guidance in *Drosophila* mushroom bodies. *Neuron* 43, 673–686. doi: 10.1016/j.neuron.2004.07.020

Zipursky, S.L., and Grueber, W.B. (2013). The molecular basis of self-avoidance. *Ann. Rev. Neurosci.* 36, 547–568. doi: 10.1146/annurev-neuro-062111-150414

Zipursky, S.L., and Sanes, J.R. (2010). Chemoaffinity revisited: dscams, protocadherins, and neural circuit assembly. *Cell* 143, 343–353. doi: 10.1016/j.cell.2010.10.009

Tables

Pcdh Fragment	Oligomeric state	Dissociation constant (μM)
$\gamma\text{A1}_{\text{EC1-3}}$	Monomer	N/A
$\gamma\text{A1}_{\text{EC1-4}}$	Dimer	13.3 ± 0.93
$\gamma\text{A4}_{\text{EC1-3}}$	Monomer	N/A
$\gamma\text{A4}_{\text{EC1-4}}$	Dimer	45.3 ± 1.52
$\gamma\text{A8}_{\text{EC1-4}}$	Dimer [†]	$30 \pm 1.5^{\dagger}$
$\gamma\text{A9}_{\text{EC1-5}}$	Dimer	8.61 ± 0.35
$\gamma\text{B5}_{\text{EC1-4}}$	Dimer	79.1 ± 4.3
$\gamma\text{B6}_{\text{EC1-3}}$	Monomer	N/A
$\gamma\text{B6}_{\text{EC1-4}}$	Dimer [†]	$29 \pm 4.9^{\dagger}$
$\gamma\text{B7}_{\text{EC1-4}}$	Dimer	146.7 ± 44.2
$\gamma\text{C3}_{\text{EC1-4}}$	Dimer	115 ± 1.49 ($K_I/K_D = 1.56$)
$\gamma\text{C5}_{\text{EC1-3}}$	Monomer [†]	N/A
$\gamma\text{C5}_{\text{EC1-5}}$	Dimer [†]	$100 \pm 4.33^{\dagger}$

Table 1. EC1–4 is required for *trans* dimerization for all γ -Pcdh subfamilies

Oligomeric state and binding affinity of N-terminal Pcdh fragments in solution determined by sedimentation equilibrium analytical ultracentrifugation. The ratio between the isodesmic constant (K_I) and dissociation constant (K_D) is given for cases where it is less than two, indicating possible non-specific binding. [†]Data from Rubinstein et al., 2015.

Pcdh Fragment	Oligomeric state	Dissociation constant (μM)
<i>Entire ectodomains</i>		
$\beta_{5\text{EC1-6}}$	Tetramer	3.93/3.19*
$\gamma\text{A1}_{\text{EC1-6}}$	Dimer	1.18 ± 0.31
$\gamma\text{A4}_{\text{EC1-6}}$	Dimer	27.8 ± 0.73
$\gamma\text{A9}_{\text{EC1-6}}$	Dimer	7.81 ± 1.05
$\gamma\text{B2}_{\text{EC1-6}}$	Tetramer	2.8/8.9*
$\gamma\text{B6}_{\text{EC1-6}}$	Tetramer	3.38/2.68*
$\alpha\text{C2}_{\text{EC1-6}}$	Tetramer [†]	8.92/0.108* [†]
$\gamma\text{C3}_{\text{EC1-6}}$	Dimer	61.6 ± 0.946 ($K_I/K_D = 1.51$)
$\gamma\text{C5}_{\text{EC1-6}}$	Tetramer [†]	18/7.64* [†]
$\alpha_{7\text{EC1-5}}/\gamma\text{C3}_{\text{EC6}}$ chimera	Tetramer	2.98/3.87*
<i>Fragments containing the cis interaction region</i>		
$\gamma\text{A1}_{\text{EC2-6}}$	Non-specific dimer	403 ± 7.74 ($K_I/K_D = 1.15$)
$\gamma\text{A4}_{\text{EC3-6}}$	Monomer	N/A
$\gamma\text{B2}_{\text{EC3-6}}$	Dimer	80.1 ± 12.8
$\alpha\text{C2}_{\text{EC2-6}}$	Dimer [†]	8.92 ± 0.28 [†]
$\gamma\text{C3}_{\text{EC3-6}}$	Monomer	N/A
$\gamma\text{C5}_{\text{EC2-6}}$	Dimer [†]	18.4 ± 0.24 [†]

Table 2. EC6-dependent homophilic ‘cis’ interactions are observed for β -, γ - and some C-type Pcdhs but not for γA -Pcdhs

Oligomeric state and binding affinity of Pcdh fragments in solution determined by sedimentation equilibrium analytical ultracentrifugation. The ratio between the isodesmic constant (K_I) and dissociation constant (K_D) is given for cases where it is less than two, indicating possible non-specific binding. * K_{DS} of monomer-to-dimer / dimer-to-tetramer transitions from fitting the data to a tetramer model. [†]Data from Rubinstein et al., 2015.

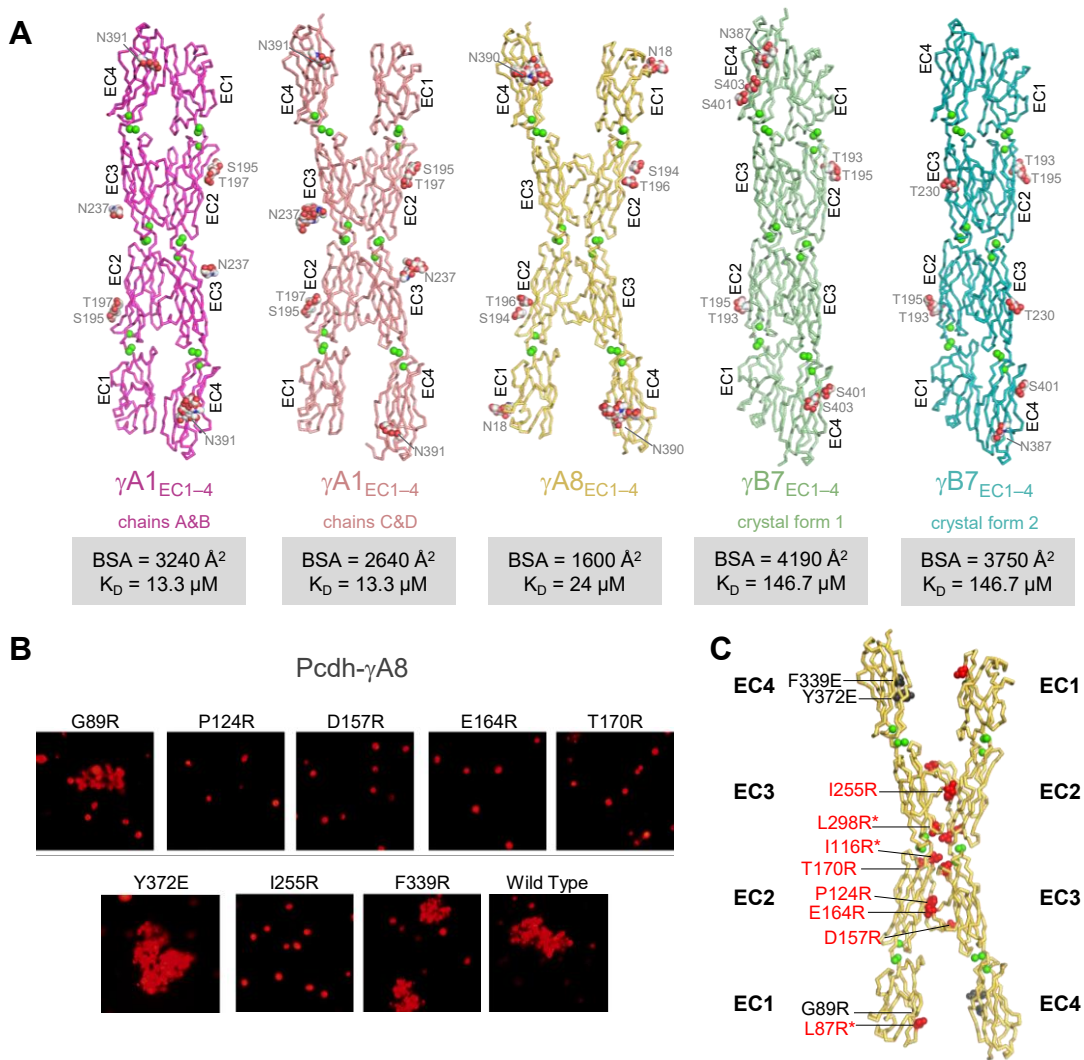


Figure 1: Crystal structures of γA and γB Pcdh cell-cell recognition dimers

A. *Trans*-dimer structures of γA1, γA8, and γB7 EC1–4 fragments. The γA1_{EC1–4} structure contained two distinct dimers in the asymmetric unit (chain A&B in magenta and chain C&D in salmon). The structures are shown in ribbon depiction with bound calcium ions shown as green spheres. Glycosylated residues are labeled, and glycans are shown as red, white and blue spheres. The buried surface area (BSA) (see Figure 1—source data 5) and the dimer dissociation constant (K_D) in solution (see Table1) are given beneath each structure.

B. K562 cell aggregation assays with γA8 mutants confirm the *trans*-dimer interface.

C. Mutations that prevent cell aggregation are shown on the γA8 dimer structure as red spheres and those which had no effect are shown as grey spheres. *Data from Rubinstein et al., 2015.

See also Figure 1—figure supplements 1–3 and source data 1–5

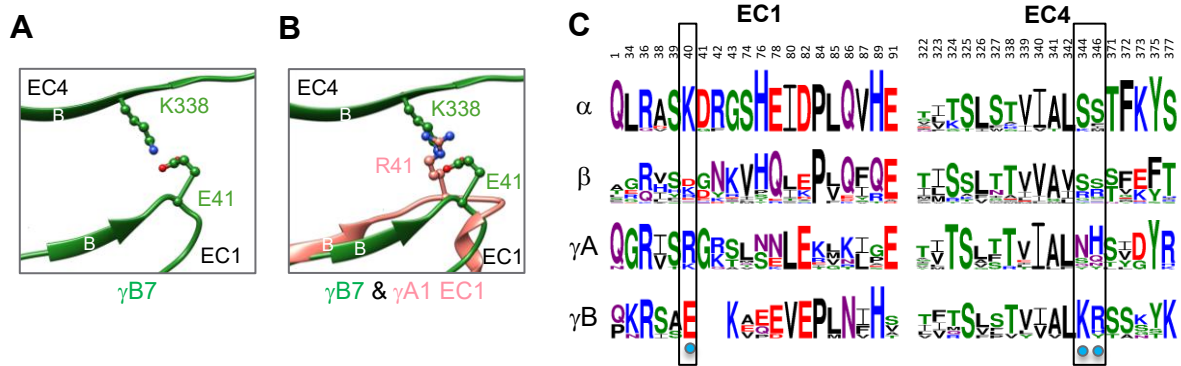


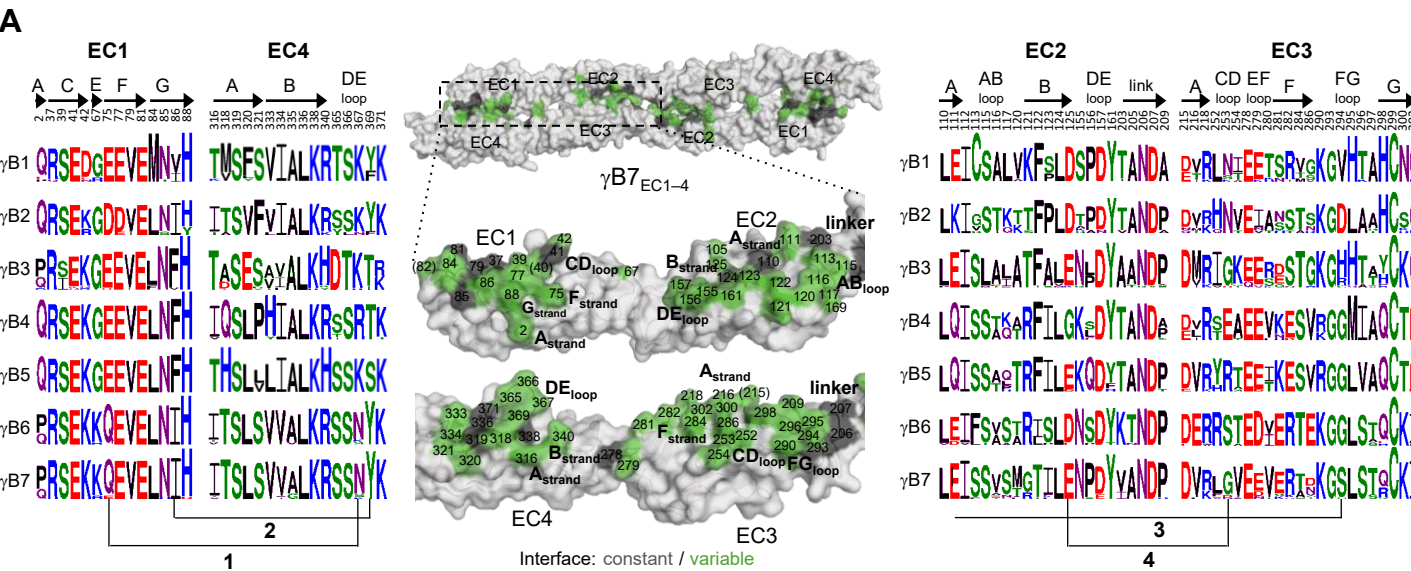
Figure 2: Pcdh interfamily specificity determinants in EC1 and EC4

A. Close-up view of the salt-bridge formed between E41 and K338 at the EC1:EC4 interface in γB7.

B. Close-up view of a structural comparison between EC1 of γA1 (salmon) and γB7 (green) structures. The interacting region in the γB7 EC4 domain is shown. Side chains are shown for residues E41, K338 of γB7 and residue γA1 R41 residue. While γB7 K338 form a salt bridge with residue E41 in the homodimer, it would likely to clash with γA1 R41 in a putative heterophilic complex.

C. Sequence logos for interfacial residues in EC1:EC4 for each of the mouse alternate isoforms (α, β, γA and γB). The logos are generated from sequence alignments of all isoform for each subfamily (see Materials and Methods). Numbering at the top of the alignment correspond to Pcdhα7 residues. The black boxes highlights the sequence positions of residues participating in the EC1:EC4 salt-bridge interaction seen in the γB7 structure (E41 and K338, γB7 numbering).

See also Figure 2—figure supplement 1



B

	Sequence positions	Residues	Isoforms	Type of interaction
1	75 _{EC1} :367 _{EC4}	E:K	γ B1–5	Electrostatic
		Q:N	γ B6–7	
2	86 _{EC1} :369 _{EC4}	I/V:Y/F	γ B1,2,6,7	Van der Waals
		F:T/S	γ B3–5	
3	111 _{EC2} :298 _{EC3}	E:S/H/R	γ B3,7	Electrostatic
		K:D	γ B2	
4	125 _{EC2} :253 _{EC3}	E:G	γ B4	Van der Waals
		G:E	γ B5	

Figure 3: γ B-Pcdh *trans*-binding specificity is encoded across the entire EC1–4 interface

A. The central panel shows a surface view of the γ B7_{EC1–4} dimer, with half of the two-fold symmetric interface opened out to reveal the interacting faces. Interfacial residues are labeled and colored grey if they are constant among all γ B isoforms or colored green if they vary among γ B isoforms. The left and right hand panels show sequence logos for interfacial residues in EC1:EC4 (left) and EC2:EC3 (right) for each of the 7 mouse γ B isoforms. The logos are generated from sequence alignments of multiple isoform-orthologs (see Materials and Methods). Secondary structure elements are annotated above the logos. The numbered connections between residue pairs correspond to the numbered rows in part (B).

B. Exemplar pairs of interacting residues that show conserved differences among a subset of γ B isoforms and may therefore contribute to specificity.

See also Figure 3—source data 1

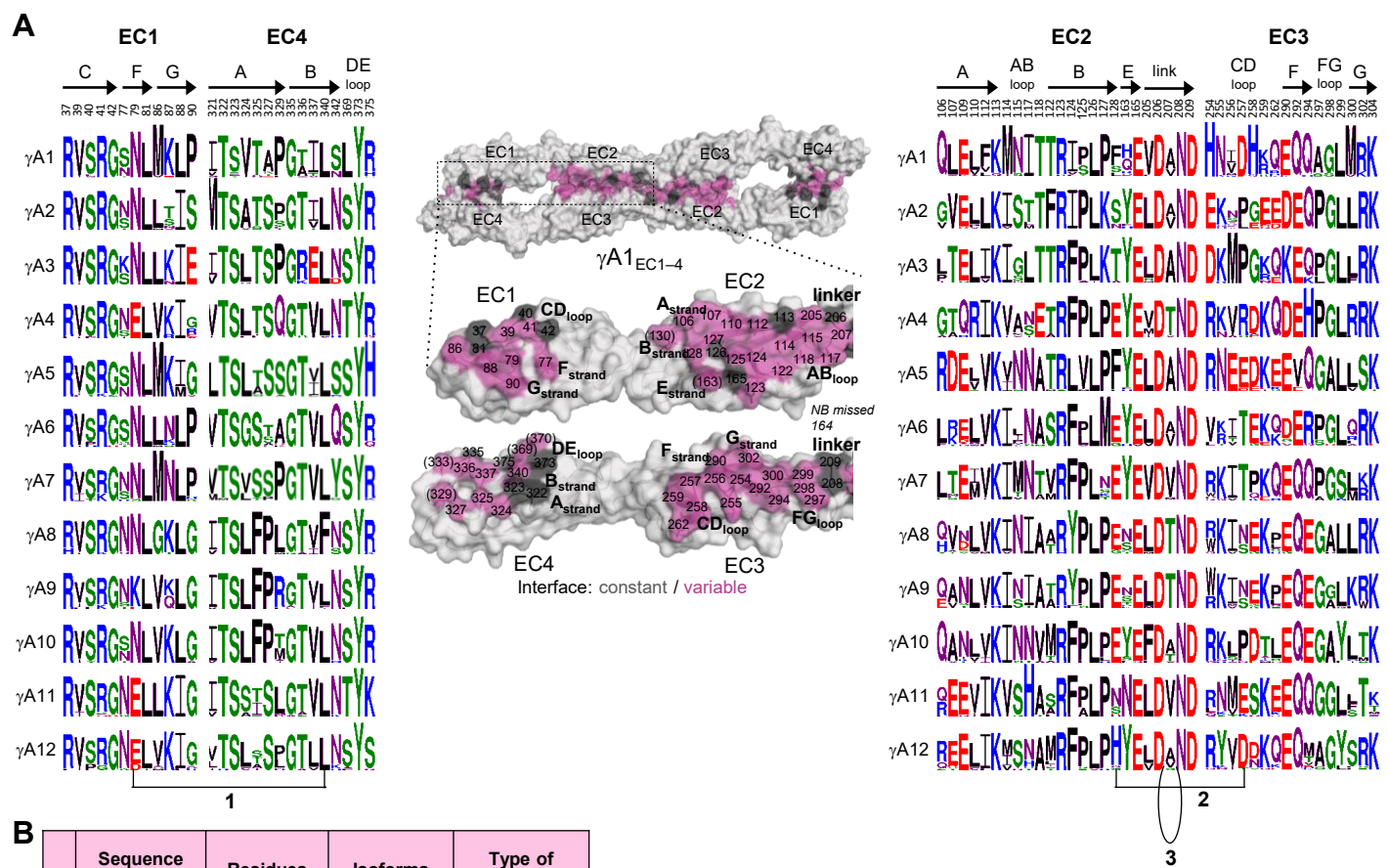


Figure 4: γ A-Pcdh *trans*-binding specificity is encoded across the entire EC1–4 interface

A. The central panel shows a surface view of the fully engaged EC1–4 γ A1 dimer, with half of the two fold symmetric interface opened out to reveal the interacting faces. Interfacial residues are labeled and colored grey if they are constant among all γ A isoforms or colored magenta if they vary among γ A isoforms. The left and right hand panels show sequence logos for interfacial residues in EC1:EC4 (left) and EC2:EC3 (right) for each of the 12 mouse γ A isoforms. The logos are generated from sequence alignments of multiple isoform-orthologs (see Materials and Methods). Secondary structure elements are annotated above the logos. The numbered connections between residue pairs correspond to the numbered rows in part (B).

B. Exemplar pairs of interacting residues that show conserved differences among a subset of γ A isoforms and may therefore contribute to specificity.

See also Figure 4—source data 1

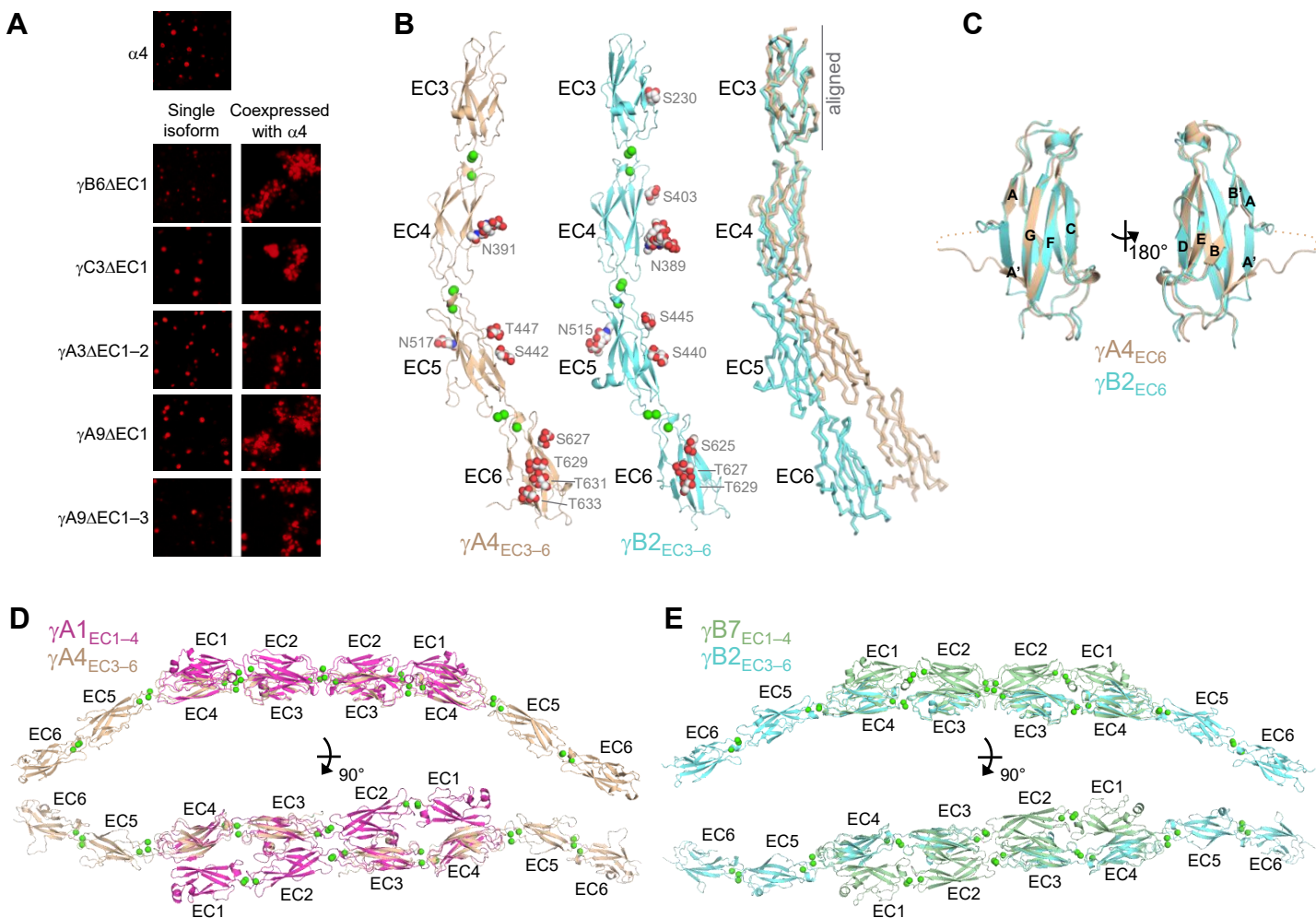


Figure 5: γA and γB Pcdhs can interact heterophilically *in cis* with α -Pcdhs and have similar *cis*-interaction region structures

A. Cell aggregation assays with K562 cells. Pcdh $\alpha 4$ cannot mediate cell aggregation when expressed alone because it does not reach the cell surface (Thu et al., 2014; top panel). Additionally γ -Pcdhs lacking part of their EC1–4 *trans* interface also cannot mediate cell aggregation (left hand panels). However when these non-adhesive fragments of γA , γB , and γC Pcdhs are co-expressed with full-length Pcdh $\alpha 4$, cell aggregation is observed (right hand panels).

B. Crystal structures of *cis* interaction containing fragments of $\gamma A4$ and $\gamma B2$. Glycosylated residues are labeled and glycans are shown as red, white and blue spheres. Bound calcium ions are shown as green spheres. Structural alignment of the EC3 domains highlights the differences in curvature between the $\gamma A4$ and $\gamma B2$ EC3–6 fragments (right panel).

C. Structural alignment of the $\gamma A4$ and $\gamma B2$ EC6 domains reveals their near identical architecture.

D. Structural alignment of the overlapping EC3–4 regions of the $\gamma A1_{EC1-4}$ dimer with the $\gamma A4_{EC3-6}$ structure provides a model for the overall architecture of γA EC1–6 dimers.

E. Structural alignment of the overlapping EC3–4 regions of the $\gamma B7_{EC1-4}$ dimer with the $\gamma B2_{EC3-6}$ structure provides a model for the overall architecture of γB EC1–6 dimers.

See also Figure 5—figure supplement 1 and source data 1

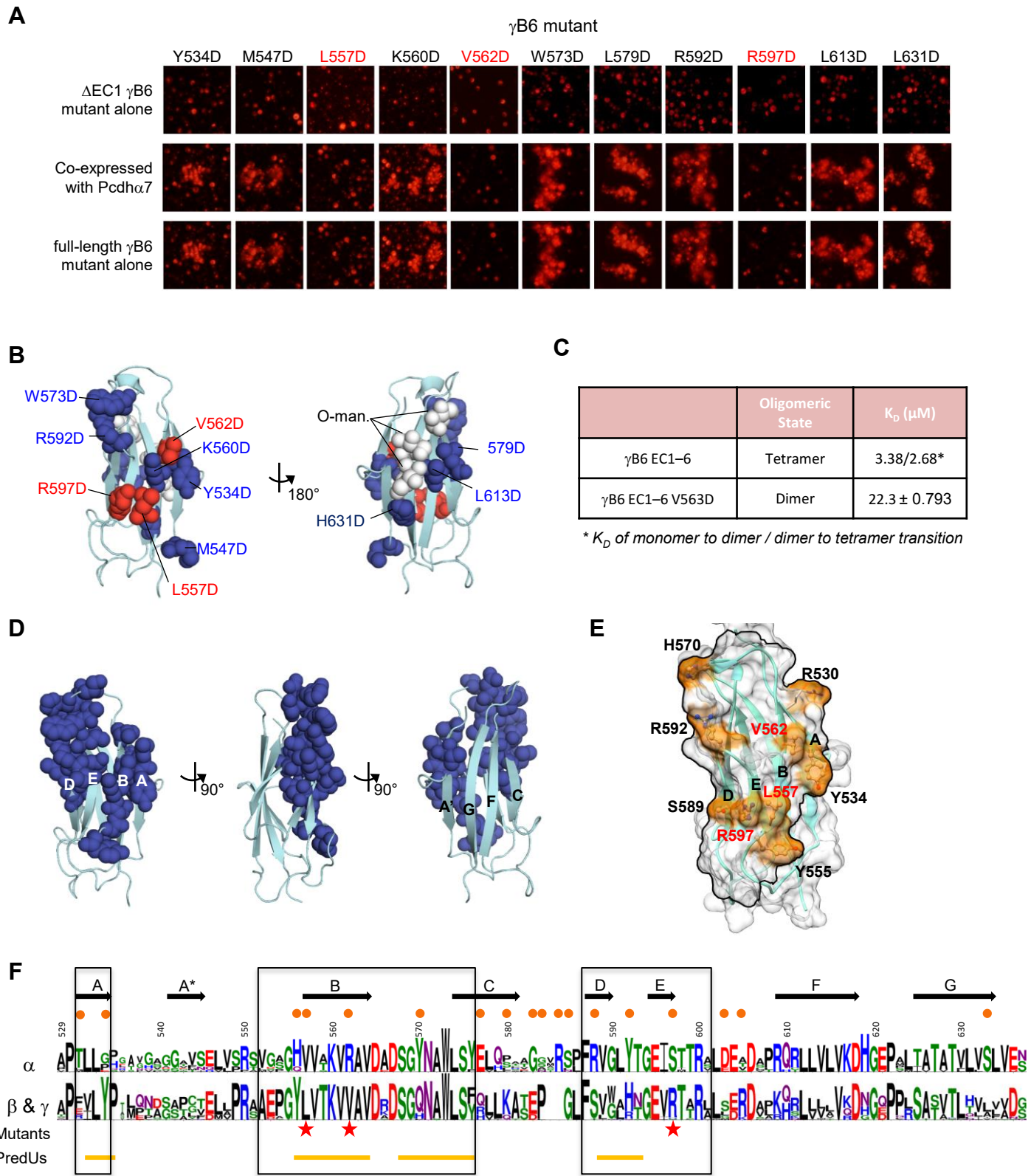


Figure 6: Identification and analysis of putative *cis* interface

A. Probing *cis* interaction interface by aspartic acid-scanning mutagenesis. Eleven point mutants of Pcdh γ B6 Δ EC1 (top) cannot mediate cell aggregation when expressed alone (top panel, γ B7 numbering). When these 11 mutants are co-expressed with a full length Pcdh α 4, cell aggregation is observed for eight of the 11 mutants (middle panels). Highlighted in red are the three mutants that cannot mediate cell aggregations likely because of failure to carry Pcdh α 4 to the cell surface. When assessed in the context of full-length Pcdh γ B6 expressed alone, the three mutants that were unable to deliver Pcdh α 4 to cell surface did not aggregate cells, while the other eight mutants did mediate cell aggregations.

B. Residues mutated to aspartic acid are drawn in space filling representation. In red are the three mutations that disrupt cell surface delivery and in blue are mutations that did not disrupt cell delivery. Glycans are shown as white spheres and are found only on one side of the molecule – the side opposite to the mutations disrupting cell delivery.

C. Analytical ultracentrifugation analysis shows that while wild type Pcdh γ B6 is a tetramer in solution, the EC6 mutant that was found to disrupt cell surface delivery (Figure 6A and B) is a dimer in solution.

D. Residues predicted to be interfacial are drawn in space filling representation. Predicted interfacial residues occupy only one side of the molecule (composed of A, B, D, and E strands). This is the same side that was indicated by the mutagenesis approach to mediate *cis* interactions and opposite to the side that contains the glycans.

E. Surface representation of the EC6 domain of γ B6 is shown. Black lines frame the face of the molecule containing mutations that disrupt cell surface delivery (labeled in red) and the predicted interface residues. Nine surface exposed residues that show different amino acid compositions between α -Pcdhs and the carrier β - and γ -Pcdhs are labeled and colored in orange.

F. Sequence logos for the EC6 domain for the α and the β and γ (γ A and γ B) Pcdhs. The logos are generated from sequence alignments of mouse α 1–12 isoforms and all mouse β , γ A and γ B isoforms. Sequence positions that differ between alphas and carrier isoforms are highlighted by orange circles (top of the logo). The three mutants that disrupt cell surface delivery are highlighted by red stars (bottom of logo). Regions predicted by PredUs to be at the interface are marked by yellow lines (bottom of the logo). Sequence positions that are part of the face likely to contain the *cis* interface are boxed. Secondary structure is shown on top.

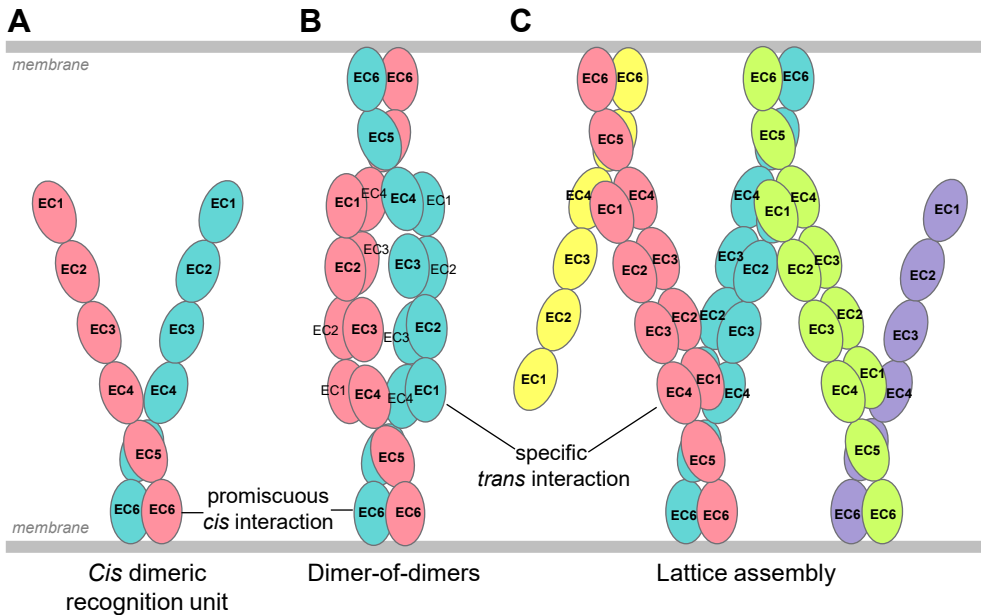


Figure 7: Schematic of possible modes of Pcdh-mediated cell-cell recognition

A. Pcdhs form homophilic and/or heterophilic *cis*-dimers, which are thought to be the cell-cell recognition unit.

B. One possible Pcdh recognition complex is a dimer-of-dimers, which has been observed in solution for homophilic complexes of β , γ B and some C-type Pcdhs (Table 2). In this model the specificity of the *trans*-interaction would require both arms of the *cis*-dimer to match for recognition (Rubinstein et al., 2015).

C. An alternative recognition complex that has been proposed is a linear zipper (Rubinstein et al., 2015). In this model only one arm of opposing *cis*-dimers needs to match to join the assembly, but incorporation of a dimer containing an isoform that is not expressed by the opposing cell would terminate growth of the intercellular Pcdh zipper.

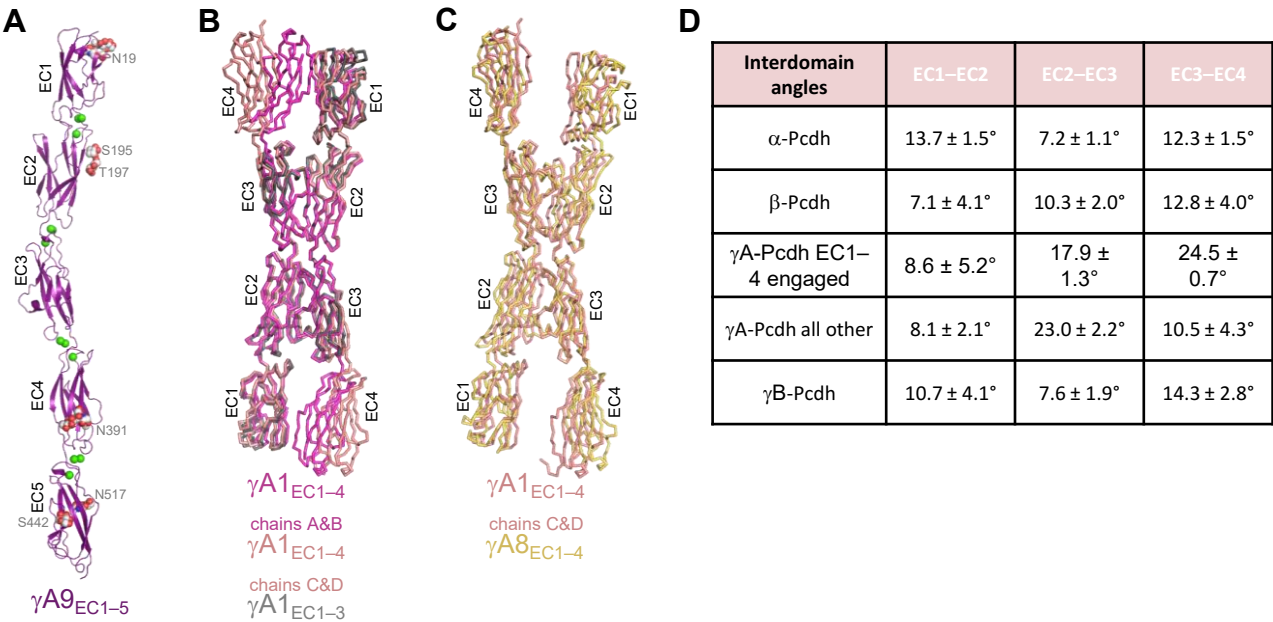


Figure 1—figure supplement 1: γ A9_{EC1–5} monomer crystal structure and γ A-Pcdh structural variability

A. γ A9_{EC1–5} crystal structure with bound calciums shown as green spheres. Glycosylated residues are labeled, and the glycans are shown as red, white and blue spheres.

B. Superposition of the two γ A1_{EC1–4} dimers in the crystal structure onto the γ A1_{EC1–3} structure (PDB: 4ZI9) highlighting the similarity of the EC2–3 dimer region.

C. Superposition of the EC2–3 engaged γ A1_{EC1–4} and γ A8_{EC1–4} dimer structures.

D. Average inter-EC domain angles for all EC1–4 containing Pcdh structures highlighting subfamily differences in architecture. Angles are given as the deviation from 180° (see schematic on the right). γ A-Pcdhs have larger EC2–EC3 angles than those from all other subfamilies. This means a larger EC3–EC4 angle is required for γ A isoformss to form a fully engaged EC1–4 dimer. Interdomain angles for each included structure are listed in Figure 1—source data 4.

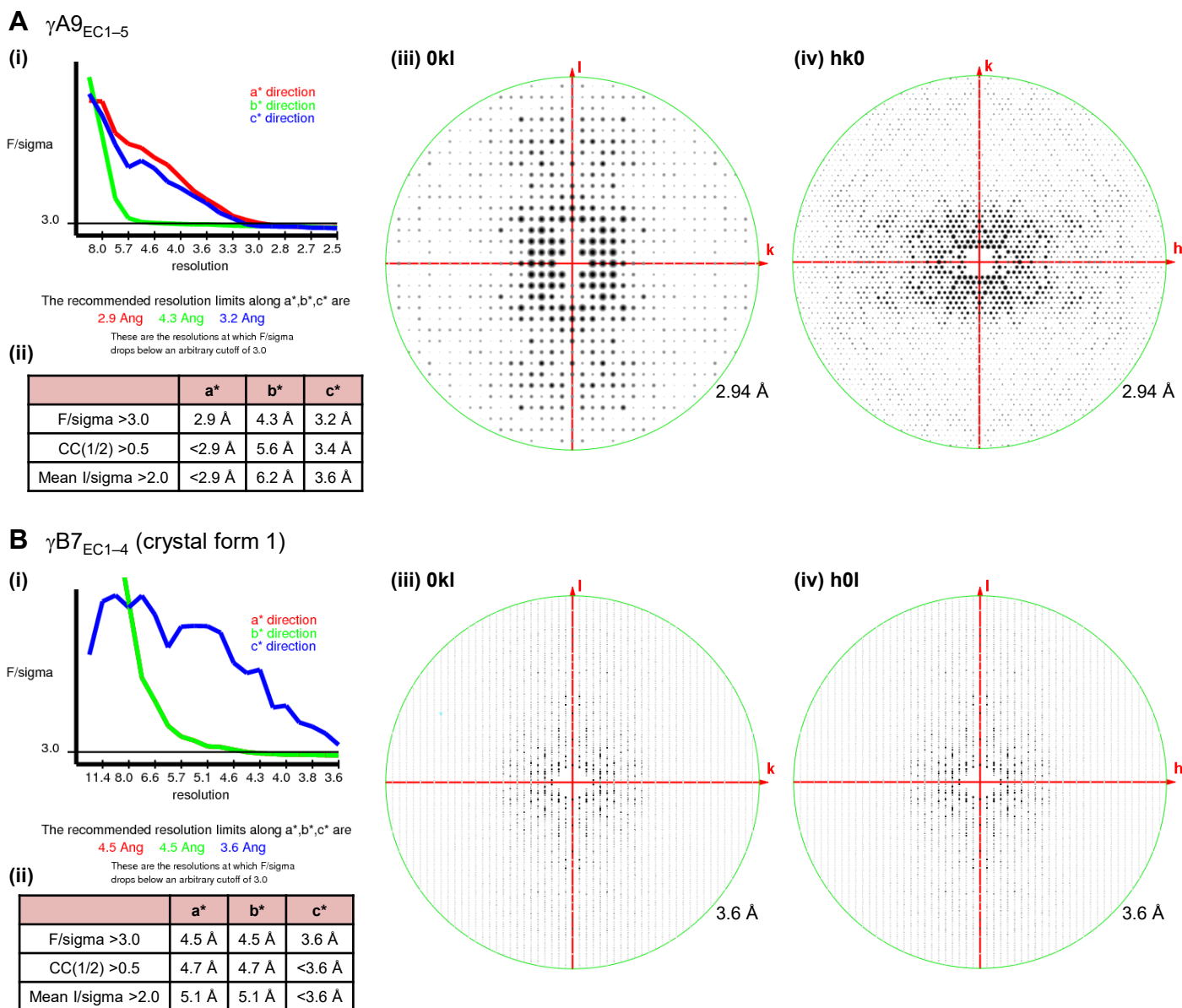


Figure 1—figure supplement 2: X-ray diffraction anisotropy of the $\gamma A9_{EC1-5}$ and $\gamma B7_{EC1-4}$ (crystal form 1) crystals

A. $\gamma A9_{EC1-5}$ crystal (i) UCLA Diffraction Anisotropy Server (Strong et al., 2006) output showing the data strength as measured by F/sigma along the a^* , b^* and c^* axes. (ii) The diffraction limits along the a^* , b^* and c^* axes determined by three different methods: F/sigma from (i), and the correlation coefficient (CC) and I/sigma limits calculated by Aimless (Evans et al., 2006; Evans and Murshudov, 2013). (iii–iv) Synthetic precession photographs of the X-ray diffraction in the $h=0$ plane (iii) and the $l=0$ plane (iv) showing the weaker diffraction along k .

B. $\gamma B7_{EC1-4}$ crystal form 1 (i) and (ii) as above. (iii–iv) Synthetic precession photographs of the X-ray diffraction in the $h=0$ plane (iii) and the $k=0$ plane (iv) showing the weaker diffraction along k and h .

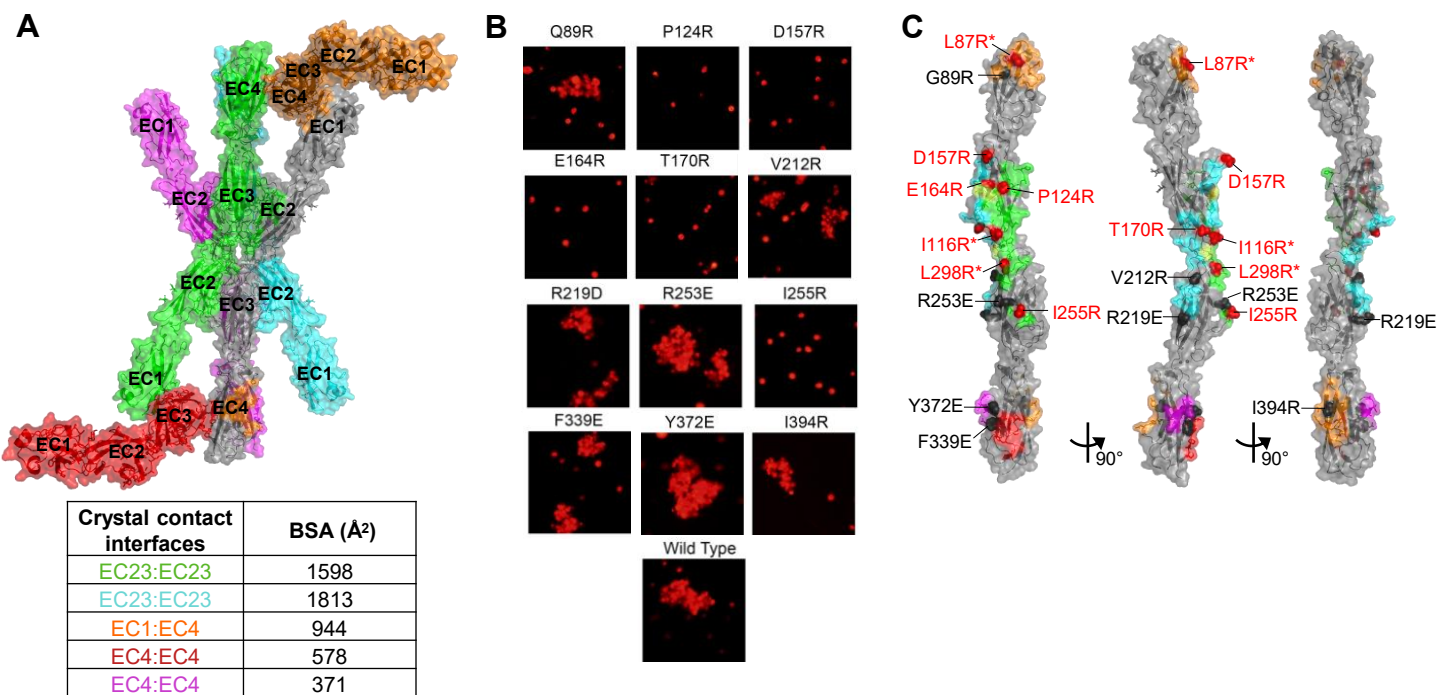


Figure 1—figure supplement 3: Mutagenesis experiments identifying the γ A8 *trans* interface among the various crystal lattice contacts

A. Surface view of the γ A8_{EC1-4} crystal structure showing the one molecule in the asymmetric unit (gray) with all the symmetry related molecules in the crystal shown. There are two anti-parallel EC2–3 contacts in the crystal (green and cyan), both have considerable buried surface areas (BSAs). The other crystal contacts are much smaller and involve EC1 and/or EC4.

B. K562 cells were transfected with Pcdh γ A8 mutants targeting the various crystal lattice contacts. Cell aggregation was observed for some of the mutants, indicating that they do not disrupt the Pcdh γ A8 recognition interface, whereas the other Pcdh γ A8 mutants failed to mediate cell aggregation, suggesting that the mutation was sufficient to disrupt the recognition interface.

C. Surface view of the γ A8_{EC1-4} structure with the interfaces mediating each of the crystal lattice contacts colored to match (A). Sites of mutations that disrupted the recognition interface are shown in red, and those that did not disrupt cell-cell recognition are shown in black. The mutations in EC2 and EC3 that interfered with the recognition interface were all in the green EC23 interface. This interacting surface matches that observed for all other Pcdh *trans* dimers. *Data from Rubinstein et al., 2015.

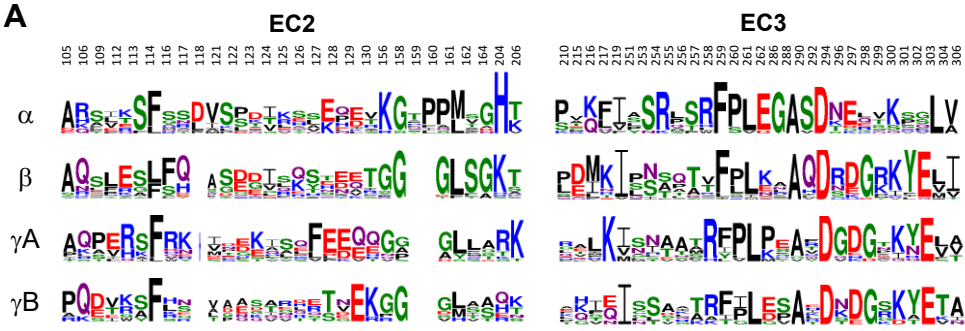


Figure 2—figure supplement 1: Sequence variability among Pcdh subfamilies in the interfacial regions of EC2 and EC3

Sequence logos for interfacial residues in EC2:EC3 for each of the mouse alternate isoforms (α , β , γA and γB). The logos are generated from sequence alignments of all isoform for each subfamily (see Materials and Methods). Numbering at the top of the alignment correspond to Pcdh $\alpha 7$ residues.

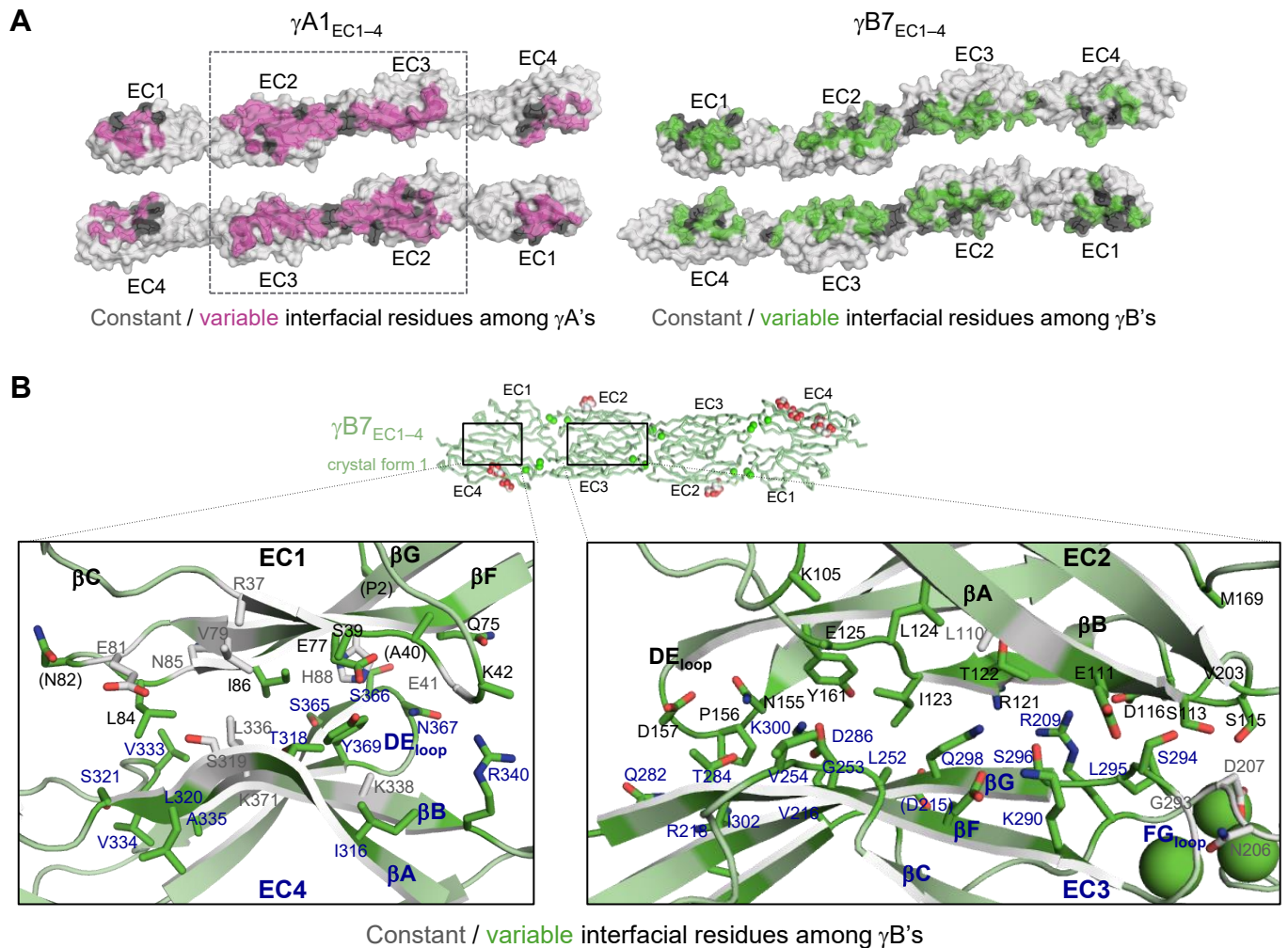
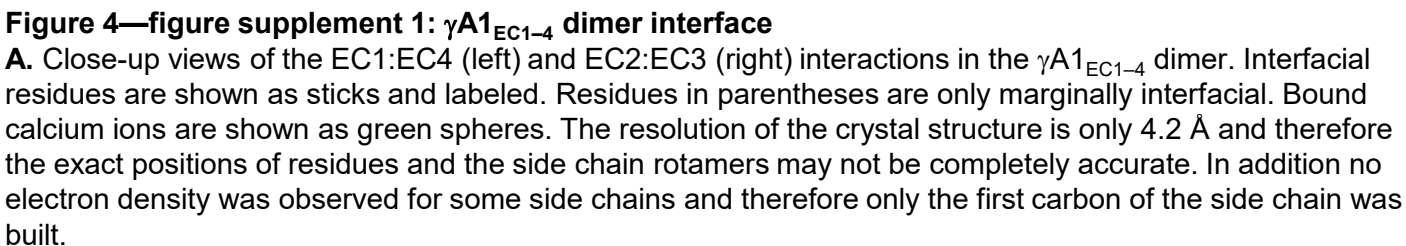


Figure 3—figure supplement 1: $\gamma B7_{EC1-4}$ dimer interface

A. Surface views of opened out $\gamma A1_{EC1-4}$ (left) and $\gamma B7_{EC1-4}$ (right) dimers with the interface residues highlighted. The majority of interfacial residues for both γA and γB Pcdhs vary among isoforms. $\gamma A1$ EC2–3 regions are boxed since these regions are interacting in all $\gamma A1$ and $\gamma A8$ crystal structures.

B. Close-up views of the EC1:EC4 (left) and EC2:EC3 (right) interactions in the $\gamma B7_{EC1-4}$ dimer. Interfacial residues are shown as sticks and labeled. Residues in parentheses are only marginally interfacial. Bound calcium ions are shown as green spheres. The resolution of the crystal structure is only 3.6 Å and therefore the exact positions of residues and the side chain rotamers may not be completely accurate.



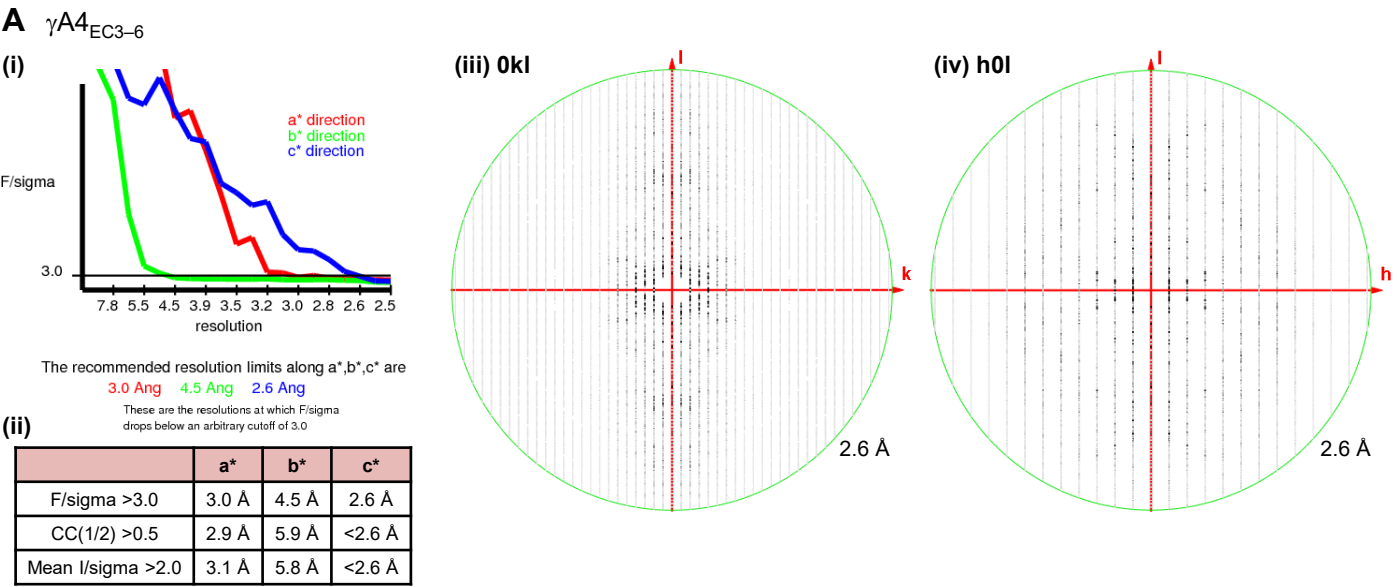
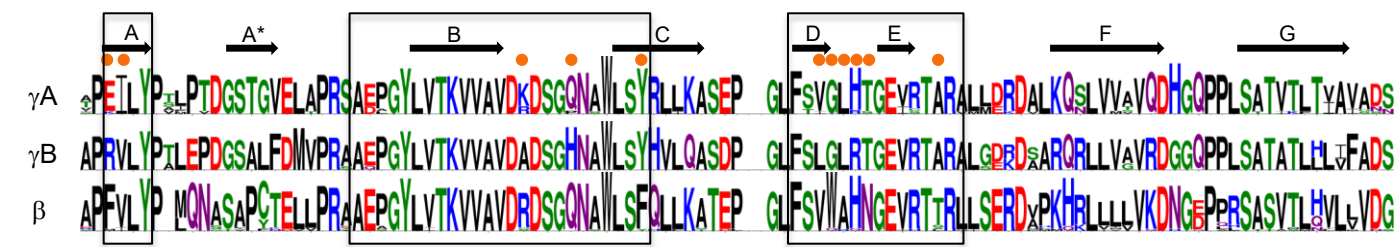


Figure 5—figure supplement 1: X-ray diffraction anisotropy of the $\gamma A4_{EC3-6}$ crystal

A. $\gamma A4_{EC3-6}$ crystal (i) UCLA Diffraction Anisotropy Server (Strong et al., 2006) output showing the data strength as measured by F/sigma along the a*, b* and c* axes. (ii) The diffraction limits along the a*, b* and c* axes determined by three different methods: F/sigma from (i), and the correlation coefficient (CC) and I/sigma limits calculated by Aimless (Evans et al., 2006; Evans and Murshudov, 2013). (iii–iv) Synthetic precession photographs of the X-ray diffraction in the h=0 plane (iii) and the k=0 plane (iv) showing the much weaker diffraction along k and slightly weaker diffraction along h.

A EC6 sequence analysis



B

	Average pairwise sequence identity in EC6
Alternate α -Pcdhs	78%
Alternate β -Pcdhs	90%
Alternate γ A-Pcdhs	90%
Alternate γ B-Pcdhs	96%
C-type Pcdhs	45%

Figure 6—figure supplement 1: EC6 sequence analysis

A. Sequence logos generated from alignments of all mouse alternate γ A, γ B, and β Pcdhs. Secondary structure elements from the γ B2 EC6 structure are annotated above. The face of EC6 that is predicted to mediate *cis* interactions from mutagenesis experiments and computational methods (Figure 6) is boxed. Residues in these regions that show conserved differences between γ A, γ B, and β Pcdhs are marked with orange dots.

B. Average pairwise amino acid sequence identities between EC6 domains of mouse Pcdh isoforms from each Pcdh sub-family.

	Pcdh γ A1 _{EC1-4}	Pcdh γ A8 _{EC1-4}	Pcdh γ A9 _{EC1-5}	Pcdh γ B7 _{EC1-4} crystal form 1	Pcdh γ B7 _{EC1-4} crystal form 2
Data collection					
Date	03/24/2016	06/14/2014	06/26/2015	10/25/2015	07/22/2016
Beamline	APS 24ID-E	APS 24ID-E	APS 24ID-E	APS 24ID-C	APS 24ID-C
Wavelength (Å)	0.97918	0.97915	0.97918	0.97930	0.97919
Space group	P3 ₁ 21	I4 ₁ 22	C2	P4 ₁ 2 ₁ 2	P2 ₁
<i>Cell dimensions</i>					
a, b, c (Å)	107.87, 107.87, 463.08	257.560, 257.560, 105.190	191.677, 107.614, 49.866	97.15, 97.15, 312.39	83.81, 45.55, 127.07
α , β , γ (°)	90, 90, 120	90, 90, 90	90, 97.14, 90	90, 90, 90	90, 96.91, 90
Resolution (Å)	40.00–4.20 (4.54–4.20)	66.51–3.60 (3.97–3.60)	38.80–2.94 (3.12–2.94)	104.13–3.59 (3.93–3.59)	39.95–3.10 (3.31–3.10)
No. of reflections	85269	146977	78312	228986	59312
Unique reflections	23885	20729	21184	18347	17677
R _{merge}	0.379 (2.646)	0.172 (0.913)	0.229 (3.156)	0.183 (3.722)	0.180 (1.512)
CC(1/2)	0.991 (0.318)	0.998 (0.864)	0.994 (0.582)	1.000 (0.741)	0.982 (0.585)
I/ σ I	3.1 (1.1)	8.6 (2.3)	4.5 (0.6)	9.0 (1.0)	5.5 (0.8)
Completeness (%)	99.9 (100.0)	99.8 (100.0)	98.9 (98.8)	99.6 (98.7)	99.5 (99.2)
Redundancy	3.6 (3.6)	7.1 (7.3)	3.7 (3.8)	12.5 (12.3)	3.4 (3.3)
Refinement					
Resolution (Å)	40–4.2	20–3.6	20–2.94/4.3/3.2	20–4.5/4.5/3.6	20–3.1
Unique reflections	23652	20598	13469	11902	17214
Completeness in diffracting sphere/ ellipsoid* (%)	99.2	99.8	92.9*	90.4*	97.2
R _{work} / R _{free} (%)	28.66 / 31.36	21.20 / 23.95	23.74 / 28.73	24.21 / 27.87	25.58 / 30.98
Molecules in ASU	4	1	1	2	2
<i>Number of residues</i>					
Protein	1659	416	523	826	818
Carbohydrate	22	8	9	10	8
Small molecule	0	0	2	0	1
Ion	36	9	15	18	18
Water	0	0	37	0	5
<i>R.m.s. deviations</i>					
Bond lengths (Å)	0.004	0.012	0.003	0.003	0.002
Bond angles (°)	0.822	0.546	0.698	0.588	0.585
<i>Ramachandran</i>					
Favored (%)	94.55	95.89	94.80	95.26	96.56
Allowed (%)	5.45	3.86	5.01	4.74	3.44
Outliers (%)	0.00	0.24	0.19	0.00	0.00
Rotamer outliers (%)	0.31	0.00	0.71	1.78	0.00
Wilson B	133.85	104.64	53.84	109.88	108.83
Overall B	196.70	157.83	97.72	235.12	81.64
PDB ID	5SZL	5SZM	5SZN	5SZO	5SZP

Figure 1—source data 1. X-ray crystallography data collection and refinement statistics

Values in parentheses are for the outer shell. ASU = asymmetric unit; R.m.s. = Root mean square.

Monomer RMSDs	$\alpha 4_{\text{EC1-4}}$	$\alpha 7_{\text{EC1-5}}$ chain A	$\beta 6_{\text{EC1-4}}$ chain A	$\beta 8_{\text{EC1-4}}$ chain A	$\gamma A1_{\text{EC1-4}}$ chain A	$\gamma A1_{\text{EC1-4}}$ chain B	$\gamma A1_{\text{EC1-4}}$ chain C	$\gamma A1_{\text{EC1-4}}$ chain D	$\gamma A8_{\text{EC1-4}}$	$\gamma A9_{\text{EC1-5}}$	$\gamma B3_{\text{EC1-4}}$	$\gamma B7_{\text{EC1-4}}$ xtal 1 chain A	$\gamma B7_{\text{EC1-4}}$ xtal 1 chain B	$\gamma B7_{\text{EC1-4}}$ xtal 2 chain A	$\gamma B7_{\text{EC1-4}}$ xtal 2 chain B
$\alpha 4_{\text{EC1-4}}$		1.5 Å (390)	2.0 Å (364)	2.0 Å (383)	3.1 Å (398)	3.1 Å (400)	3.6 Å (404)	5.5 Å (401)	6.0 Å (407)	5.7 Å (403)	XX	1.3 Å (384)	1.9 Å (391)	1.3 Å (361)	1.5 Å (383)
$\alpha 7_{\text{EC1-5}}$ chain A	1.5 Å (390)		2.0 Å (381)	2.1 Å (382)	3.4 Å (402)	3.3 Å (401)	3.4 Å (411)	5.1 Å (405)	5.4 Å (404)	5.5 Å (507)	XX	1.8 Å (372)	2.2 Å (395)	1.3 Å (345)	1.9 Å (389)
$\beta 6_{\text{EC1-4}}$ chain A	2.0 Å (364)	2.0 Å (381)		1.7 Å (367)	3.5 Å (406)	3.5 Å (400)	3.6 Å (393)	4.7 Å (375)	4.9 Å (373)	4.6 Å (366)	XX	2.4 Å (371)	2.8 Å (394)	2.1 Å (363)	2.3 Å (384)
$\beta 8_{\text{EC1-4}}$ chain A	2.0 Å (383)	2.1 Å (382)	1.7 Å (367)		3.4 Å (404)	3.0 Å (399)	3.7 Å (407)	5.3 Å (404)	4.9 Å (390)	4.9 Å (390)	XX	2.8 Å (398)	3.1 Å (394)	2.2 Å (388)	2.7 Å (389)
$\gamma A1_{\text{EC1-4}}$ chain A	3.1 Å (398)	3.4 Å (402)	3.5 Å (406)	3.4 Å (404)		0.9 Å (389)	2.9 Å (398)	4.9 Å (402)	5.0 Å (372)	5.8 Å (390)	XX	3.1 Å (401)	3.3 Å (400)	2.8 Å (397)	3.4 Å (400)
$\gamma A1_{\text{EC1-4}}$ chain B	3.1 Å (400)	3.3 Å (401)	3.5 Å (400)	3.0 Å (399)	0.9 Å (389)		2.4 Å (375)	3.2 Å (348)	3.9 Å (348)	5.8 Å (400)	XX	3.2 Å (398)	3.4 Å (400)	2.7 Å (395)	2.7 Å (389)
$\gamma A1_{\text{EC1-4}}$ chain C	3.6 Å (404)	3.4 Å (411)	3.6 Å (393)	3.7 Å (407)	2.9 Å (398)	2.4 Å (375)		1.9 Å (404)	2.0 Å (335)	3.6 Å (398)	XX	3.8 Å (407)	3.1 Å (405)	2.6 Å (392)	3.6 Å (403)
$\gamma A1_{\text{EC1-4}}$ chain D	5.5 Å (401)	5.1 Å (405)	4.7 Å (375)	5.3 Å (404)	4.9 Å (402)	3.2 Å (348)	1.9 Å (404)		1.2 Å (336)	2.0 Å (345)	XX	6.0 Å (411)	5.1 Å (407)	4.6 Å (402)	5.4 Å (400)
$\gamma A8_{\text{EC1-4}}$	6.0 Å (407)	5.4 Å (404)	4.9 Å (373)	5.6 Å (387)	5.0 Å (372)	3.9 Å (348)	2.0 Å (335)	1.2 Å (336)		2.2 Å (342)	XX	6.2 Å (408)	5.0 Å (401)	5.0 Å (403)	5.4 Å (396)
$\gamma A9_{\text{EC1-5}}$	5.7 Å (403)	5.5 Å (507)	4.6 Å (366)	4.9 Å (390)	5.8 Å (390)	5.8 Å (400)	3.6 Å (398)	2.0 Å (345)	2.2 Å (342)		XX	6.0 Å (391)	5.3 Å (391)	4.8 Å (388)	5.4 Å (388)
$\gamma B3_{\text{EC1-4}}$	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX		XX	XX	XX	XX
$\gamma B7_{\text{EC1-4}}$ xtal 1 chA	1.3 Å (384)	1.8 Å (372)	2.4 Å (371)	2.8 Å (398)	3.1 Å (401)	3.2 Å (398)	3.8 Å (407)	6.0 Å (411)	6.2 Å (408)	6.0 Å (391)	XX		1.3 Å (398)	1.5 Å (408)	1.2 Å (396)
$\gamma B7_{\text{EC1-4}}$ xtal 1 chB	1.9 Å (391)	2.2 Å (395)	2.8 Å (394)	3.1 Å (394)	3.3 Å (400)	3.4 Å (400)	3.1 Å (405)	5.1 Å (407)	5.0 Å (401)	5.3 Å (391)	XX	1.3 Å (398)		1.4 Å (398)	1.0 Å (380)
$\gamma B7_{\text{EC1-4}}$ xtal 2 chA	1.3 Å (361)	1.3 Å (345)	2.1 Å (363)	2.2 Å (388)	2.8 Å (397)	2.7 Å (395)	2.6 Å (392)	4.6 Å (402)	5.0 Å (403)	4.8 Å (388)	XX	1.5 Å (408)	1.4 Å (398)		1.4 Å (383)
$\gamma B7_{\text{EC1-4}}$ xtal 2 chB	1.5 Å (383)	1.9 Å (389)	2.3 Å (384)	2.7 Å (389)	3.4 Å (400)	2.7 Å (389)	3.6 Å (403)	5.4 Å (400)	5.4 Å (396)	5.4 Å (388)	XX	1.2 Å (396)	1.0 Å (380)	1.4 Å (383)	

Figure 1—source data 2. Overall structural similarity between EC1–4 regions of α -, β -, and γ -Pcdh structures

Root mean square deviations over aligned C α 's (RMSDs) between pairs of Pcdh individual protomers. xtal 1 = crystal form 1; xtal 2 = crystal form 2. The $\alpha 4_{\text{EC1-4}}$, $\alpha 7_{\text{EC1-5}}$, $\beta 6_{\text{EC1-4}}$, $\beta 8_{\text{EC1-4}}$, and $\gamma B3_{\text{EC1-4}}$ structures correspond to PDBs: 5DZW, 5DZV, 5DZX, 5DZY, and 5K8R. *The RMSDs for $\gamma B3_{\text{EC1-4}}$ will be included when the structure coordinates are released in the PDB.*

Dimer RMSDs	$\alpha 4$	$\alpha 7$	$\beta 6$	$\beta 8$ (chains A&B)	$\gamma A1$ (chains A&B)	$\gamma A1$ (chains C&D)	$\gamma A8$	$\gamma B3$	$\gamma B7$ (crystal form 1)	$\gamma B7$ (crystal form 2)
$\alpha 4$		1.9 Å (779)	4.7 Å (788)	3.7 Å (776)	4.7 Å (803)	7.7 Å (782)	11.7 Å (812)	XX	2.2 Å (775)	1.8 Å (756)
$\alpha 7$	1.9 Å (779)		4.7 Å (793)	3.9 Å (783)	4.3 Å (796)	6.9 Å (811)	10.3 Å (812)	XX	2.9 Å (798)	2.5 Å (775)
$\beta 6$	4.7 Å (788)	4.7 Å (793)		1.6 Å (715)	5.0 Å (814)	7.9 Å (804)	11.2 Å (793)	XX	4.0 Å (800)	3.9 Å (791)
$\beta 8$ (chains A&B)	3.7 Å (776)	3.9 Å (783)	1.6 Å (715)		4.5 Å (810)	7.2 Å (799)	10.2 Å (786)	XX	3.3 Å (784)	3.3 Å (790)
$\gamma A1$ (chains A&B)	4.7 Å (803)	4.3 Å (796)	5.0 Å (814)	4.5 Å (810)		5.6 Å (791)	9.2 Å (792)	XX	4.7 Å (804)	4.5 Å (800)
$\gamma A1$ (chains C&D)	7.7 Å (782)	6.9 Å (811)	7.9 Å (804)	7.2 Å (799)	5.6 Å (791)		4.8 Å (802)	XX	7.7 Å (802)	7.3 Å (776)
$\gamma A8$	11.7 Å (812)	10.3 Å (812)	11.2 Å (793)	10.2 Å (786)	9.2 Å (792)	4.8 Å (802)		XX	11.2 Å (800)	10.9 Å (795)
$\gamma B3$	XX	XX	XX	XX	XX	XX	XX		XX	XX
$\gamma B7$ (crystal form 1)	2.2 Å (775)	2.9 Å (798)	4.0 Å (800)	3.3 Å (784)	4.7 Å (804)	7.7 Å (802)	11.2 Å (800)	XX		1.5 Å (805)
$\gamma B7$ (crystal form 2)	1.8 Å (756)	2.5 Å (775)	3.9 Å (791)	3.3 Å (790)	4.5 Å (800)	7.3 Å (776)	10.9 Å (795)	XX	1.5 Å (805)	

Figure 1—source data 3. Overall structural similarity between α -, β -, and γ -Pcdh EC1–4 *trans* dimer structures

Root mean square deviations over aligned C α 's (RMSDs) between pairs of Pcdh *trans* dimer structures are shown. The number of aligned C α 's for each pair is given in parentheses. The $\alpha 4_{\text{EC1-4}}$, $\alpha 7_{\text{EC1-5}}$, $\beta 6_{\text{EC1-4}}$, $\beta 8_{\text{EC1-4}}$, and $\gamma B3_{\text{EC1-4}}$ structures correspond to PDBs: 5DZW, 5DZV, 5DZX, 5DZY, and 5K8R. *The RMSDs for $\gamma B3_{\text{EC1-4}}$ will be included when the structure coordinates are released in the PDB.*

Interdomain angles	EC1–EC2 (°)	EC2–EC3 (°)	EC3–EC4 (°)	EC4–EC5 (°)
$\alpha 4_{\text{EC1-4}}$	12.0	7.0	14.0	
$\alpha 7_{\text{EC1-5}}$ chain A	14.9	6.3	11.4	21.9
$\alpha 7_{\text{EC1-5}}$ chain B	14.2	8.4	11.5	22.5
<i>Average α-Pcdh</i>	13.7 ± 1.5	7.2 ± 1.1	12.3 ± 1.5	22.2 ± 0.4
$\beta 6_{\text{EC1-4}}$ chain A	13.3	11.0	8.6	
$\beta 6_{\text{EC1-4}}$ chain B	13.5	10.4	6.3	
$\beta 8_{\text{EC1-4}}$ chain A	6.0	10.1	18.2	
$\beta 8_{\text{EC1-4}}$ chain B	5.0	9.1	10.4	
$\beta 8_{\text{EC1-4}}$ chain C	4.7	7.1	14.9	
$\beta 8_{\text{EC1-4}}$ chain D	2.3	8.8	15.1	
$\beta 8_{\text{EC1-4}}$ chain E	7.0	11.6	15.4	
$\beta 8_{\text{EC1-4}}$ chain F	4.6	13.9	13.5	
<i>Average β-Pcdh</i>	7.1 ± 4.1	10.3 ± 2.0	12.8 ± 4.0	
$\gamma A1_{\text{EC1-4}}$ chain A	12.3	18.8	24.0	
$\gamma A1_{\text{EC1-4}}$ chain B	4.9	17.0	25.0	
<i>Average EC1–4 engaged γA-Pcdh</i>	8.6 ± 5.2	17.9 ± 1.3	24.5 ± 0.7	
$\gamma A1_{\text{EC1-4}}$ chain C	8.4	21.6	16.8	
$\gamma A1_{\text{EC1-4}}$ chain D	5.8	22.3	7.9	
$\gamma A8_{\text{EC1-4}}$	7.3	26.3	9.2	
$\gamma A9_{\text{EC1-5}}$	10.8	21.9	8.1	17.7
<i>Average not fully engaged γA-Pcdh</i>	8.1 ± 2.1	23.0 ± 2.2	10.5 ± 4.2	
$\gamma B3_{\text{EC1-4}}$	XX	XX	XX	
$\gamma B7_{\text{EC1-4}}$ xtal 1 chain A	12.0	6.1	15.2	
$\gamma B7_{\text{EC1-4}}$ xtal 1 chain B	11.4	10.0	15.7	
$\gamma B7_{\text{EC1-4}}$ xtal 2 chain A	4.8	6.0	10.2	
$\gamma B7_{\text{EC1-4}}$ xtal 2 chain B	14.4	8.4	16.1	
<i>Average γB-Pcdh</i>	10.7 ± 4.1	7.6 ± 1.9	14.3 ± 2.8	

Figure 1—source data 4. Interdomain angles

Interdomain angles between consecutive EC domains given as the deviation from 180°, were calculated using UCSF chimera. The $\alpha 4_{\text{EC1-4}}$, $\alpha 7_{\text{EC1-5}}$, $\beta 6_{\text{EC1-4}}$, $\beta 8_{\text{EC1-4}}$, and $\gamma B3_{\text{EC1-4}}$ structures correspond to PDBs: 5DZW, 5DZV, 5DZX, 5DZY, and 5K8R. The RMSDs for $\gamma B3_{\text{EC1-4}}$ will be included when the structure coordinates are released in the PDB.

Buried surface area (Å ²)	$\alpha 4$	$\alpha 7$	$\beta 6$	$\beta 8$ (chains A&B)	$\gamma A1$ (chains A&B)	$\gamma A1$ (chains C&D)	$\gamma A8$	$\gamma B3$	$\gamma B7$ (crystal form 1)	$\gamma B7$ (crystal form 2)
Entire interface in crystal structure	4319	3316	4554	4821	3237	2641	1598	XX	4190	3747
Entire interface including all side chains	4995	3904	4678	5093	3522	2703	1658	XX	4601	4456
EC2:EC3 interface including all side chains	2922	1975	2476	2971	2527	2703	1658	XX	2546	2846
EC1:EC4 interfaces including all side chains	1948	1929	2202	2102	997	0	0	XX	1987	1621

Figure 1—source data 5. *Trans*-dimer buried surface areas in all Pcdh EC1–4 containing crystal structures

Interfacial buried surface areas (BSAs) are given as the difference in accessible surface area over both protomers upon dimer formation. BSAs were determined using the PISA server. Unmodeled side chains in the crystal structures were generated using the Dunbrack rotamer library in UCSF chimera. The $\alpha 4_{EC1-4}$, $\alpha 7_{EC1-5}$, $\beta 6_{EC1-4}$, $\beta 8_{EC1-4}$, and $\gamma B3_{EC1-4}$ structures correspond to PDBs: 5DZW, 5DZV, 5DZX, 5DZY, and 5K8R. *The RMSDs for $\gamma B3_{EC1-4}$ will be included when the structure coordinates are released in the PDB.*

Isoform	Number of orthologs	Species
gB1	12	Mus musculus, Oryctolagus cuniculus, Rhinopithecus roxellana, Mandrillus leucophaeus, Macaca mulatta, Macaca fascicularis, Pongo abelii, Pan troglodytes, Pan paniscus, Gorilla gorilla gorilla, Physeter catodon, Camelus dromedarius, Leptonychotes weddellii
gB2	7	Monodelphis domestica, Chrysochloris asiatica, Fukomys damarensis, Bison bison bison, Leptonychotes weddellii, Rhinopithecus roxellana, Mandrillus leucophaeus, Pongo abelii
gB3	18	Oryctolagus cuniculus, Echinops telfairi, Gorilla gorilla gorilla, Pongo abelii, Rhinopithecus roxellana, Chlorocebus sabaeus, Mandrillus leucophaeus, Papio anubis, Macaca fascicularis, Macaca mulatta, Aotus nancymae, Callithrix jacchus, Leptonychotes weddellii, Chrysochloris asiatica, Trichechus manatus latirostris, Camelus dromedarius, Physeter catodon, Tursiops truncatus, Lipotes vexillifer
gB4	12	Chrysochloris asiatica, Fukomys damarensis, Callithrix jacchus, Mandrillus leucophaeus, Pongo abelii, Nomascus leucogenys, Homo sapiens, Gorilla gorilla gorilla, Camelus dromedarius, Bubalus bubalis, Physeter catodon, Lipotes vexillifer, Tursiops truncatus
gB5	9	Fukomys damarensis, Chrysochloris asiatica, Mandrillus leucophaeus, Pongo abelii, Gorilla gorilla gorilla, Homo sapiens, Lipotes vexillifer, Physeter catodon, Pantholops hodgsonii, Bison bison bison
gB6	12	Mus musculus, Fukomys damarensis, Pongo abelii, Colobus angolensis palliatus, Macaca nemestrina, Nomascus leucogenys, Gorilla gorilla gorilla, Homo sapiens, Bubalus bubalis, Pantholops hodgsonii, Lipotes vexillifer, Orcinus orca, Tursiops truncatus
gB7	19	Mus musculus, Mesocricetus auratus, Peromyscus maniculatus bairdii, Octodon degus, Tupaia chinensis, Orcinus orca, Microcebus murinus, Bubalus bubalis, Pantholops hodgsonii, Lipotes vexillifer, Tursiops truncatus, Balaenoptera acutorostrata scammoni, Galeorhinus galeus, Colobus angolensis palliatus, Pongo abelii, Homo sapiens, Gorilla gorilla gorilla, Pan paniscus, Elephantulus edwardii, Orycteropus afer

Figure 3—source data 1. List of species used in generating the sequence logo for γ B-Pcdh isoforms.

Isoform	Number of orthologs	Species
γ A1	24	Mus musculus, Rattus norvegicus, Dipodomys ordii, Jaculus jaculus, Heterocephalus glaber, Cavia porcellus, Octodon degus, Oryctolagus cuniculus, Elephantulus edwardii, Loxodonta africana, Echinops telfairi, Chrysochloris asiatica, Tarsius syrichta, Gorilla gorilla gorilla, Homo sapiens, Pan troglodytes, Pan paniscus, Colobus angolensis palliatus, Macaca nemestrina, Leptonychotes weddellii, Bubalus bubalis, Phryseter catodon, Lipotes vexillifer, Tursiops truncatus
γ A2	17	Jaculus jaculus, Mus musculus, Rattus norvegicus, Fukomys damarensis, Octodon degus, Chrysochloris asiatica, Leptonychotes weddellii, Bubalus bubalis, Phryseter catodon, Lipotes vexillifer, Orcinus orca, Tursiops truncatus, Galeopterus variegatus, Callithrix jacchus, Gorilla gorilla gorilla, Pan paniscus, Pan troglodytes,
γ A3	16	Chrysochloris asiatica, Echinops telfairi, Elephantulus edwardii, Homo sapiens, Pan troglodytes, Octodon degus, Fukomys damarensis, Galeopterus variegatus, Odobenus rosmarus divergens, Bubalus bubalis, Phryseter catodon, Lipotes vexillifer, Orcinus orca, Tursiops truncatus, Jaculus jaculus, Mus musculus, Rattus norvegicus
γ A4	14	Fukomys damarensis, Octodon degus, Mus musculus, Jaculus jaculus, Dipodomys ordii, Sorex araneus, Rhinopithecus roxellana, Homo sapiens, Gorilla gorilla gorilla, Pan troglodytes, Tarsius syrichta, Leptonychotes weddellii, Odobenus rosmarus divergens, Phryseter catodon, Bubalus bubalis
γ A5	21	Ochotona princeps, Trichechus manatus latirostris, Chrysochloris asiatica, Sorex araneus, Tarsius syrichta, Leptonychotes weddellii, Odobenus rosmarus divergens, Camelus dromedarius, Phryseter catodon, Lipotes vexillifer, Orcinus orca, Callithrix jacchus, Colobus angolensis palliatus, Pan troglodytes, Gorilla gorilla gorilla, Homo sapiens, Oryzomys latipes, Propithecus coquereli, Fukomys damarensis, Octodon degus, Jaculus jaculus, Mus musculus
γ A6	15	Sorex araneus, Dipodomys ordii, Mus musculus, Cricetulus griseus, Nannospalax galili, Jaculus jaculus, Fukomys damarensis, Octodon degus, Leptonychotes weddellii, Odobenus rosmarus divergens, Tursiops truncatus, Lipotes vexillifer, Phryseter catodon, Tarsius syrichta, Pan troglodytes, Gorilla gorilla gorilla
γ A7	14	Jaculus jaculus, Mus musculus, Cricetulus griseus, Octodon degus, Echinops telfairi, Tarsius syrichta, Pan troglodytes, Pan paniscus, Nomascus leucogenys, Sorex araneus, Leptonychotes weddellii, Odobenus rosmarus divergens, Bubalus bubalis, Lipotes vexillifer, Phryseter catodon
γ A8	14	Elephantulus edwardii, Chrysochloris asiatica, Homo sapiens, Gorilla gorilla gorilla, Leptonychotes weddellii, Camelus dromedarius, Phryseter catodon, Lipotes vexillifer, Tursiops truncatus, Octodon degus, Jaculus jaculus, Dipodomys ordii, Nannospalax galili, Mus musculus, Cricetulus griseus
γ A9	12	Leptonychotes weddellii, Elephantulus edwardii, Octodon degus, Jaculus jaculus, Mus musculus, Cricetulus griseus, Oryzomys latipes, Sorex araneus, Bison bison, Bubalus bubalis, Pantholops hodgsonii, Homo sapiens, Trichechus manatus latirostris
γ A10	16	Monodelphis domestica, Mus musculus, Elephantulus edwardii, Jaculus jaculus, Chrysochloris asiatica, Sorex araneus, Octodon degus, Ictidomys tridecemlineatus, Propithecus coquereli, Colobus angolensis palliatus, Gorilla gorilla gorilla, Leptonychotes weddellii, Odobenus rosmarus divergens, Lipotes vexillifer, Bubalus bubalis, Pantholops hodgsonii
γ A11	20	Ochotona princeps, Echinops telfairi, Chrysochloris asiatica, Mus musculus, Jaculus jaculus, Octodon degus, Ictidomys tridecemlineatus, Galeopterus variegatus, Leptonychotes weddellii, Aotus nancymaeae, Colobus angolensis palliatus, Rhinopithecus roxellana, Gorilla gorilla gorilla, Homo sapiens, Pan troglodytes, Bubalus bubalis, Balaenoptera acutorostrata scammoni, Lipotes vexillifer, Orcinus orca, Tursiops truncatus
γ A12	18	Chelonia mydas, Monodelphis domestica, Ochotona princeps, Ictidomys tridecemlineatus, Jaculus jaculus, Nannospalax galili, Chrysochloris asiatica, Elephantulus edwardii, Octodon degus, Nomascus leucogenys, Gorilla gorilla gorilla, Cercopithecus atys, Colobus angolensis palliatus, Eptesicus fuscus, Leptonychotes weddellii, Odobenus rosmarus divergens, Bubalus bubalis, Lipotes vexillifer

Figure 4—source data 1. List of species used in generating the sequence logo for γ A-Pcdh isoforms.

	Pcdh γ4_{EC3-6}	Pcdh γB2_{EC3-6}
Data collection		
Date	06/29/2016	06/29/2016
Beamline	APS 24ID-C	APS 24ID-C
Wavelength (Å)	0.97919	0.97919
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 4 ₁ 2 ₁ 2
<i>Cell dimensions</i>		
a, b, c (Å)	31.91, 63.79, 345.60	104.75, 104.75, 352.14
α , β , γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	172.16–2.56 (2.67–2.56)	39.08–2.30 (2.34–2.30)
No. of reflections	98304	485198
Unique reflections	23763	87920
R _{merge}	0.112 (3.118)	0.119 (1.886)
CC(1/2)	0.998 (0.434)	0.998 (0.882)
I/ σ I	6.6 (0.3)	6.0 (0.8)
Completeness (%)	99.3 (97.5)	99.8 (99.9)
Redundancy	4.1 (3.3)	5.5 (5.6)
Refinement		
Resolution (Å)	20–3.0/4.5/2.6	20–2.3
Unique reflections	11653	86457
Completeness in diffracting sphere/ ellipsoid* (%)	99.0*	98.3
R _{work} / R _{free} (%)	25.74 / 28.31	24.78 / 27.78
Molecules in ASU	1	3
<i>Number of residues</i>		
Protein	424	1261
Carbohydrate	9	38
Small molecule	0	5
Ion	9	27
Water	0	276
<i>R.m.s. deviations</i>		
Bond lengths (Å)	0.003	0.015
Bond angles (°)	0.655	0.920
<i>Ramachandran</i>		
Favored (%)	95.93	97.36
Allowed (%)	4.07	2.64
Outliers (%)	0.00	0.00
Rotamer outliers (%)	2.48	2.74
Wilson B	48.68	45.36
Overall B	103.26	73.95
PDB ID	5SZQ	5SZR

Figure 5—source data 1. X-ray crystallography data collection and refinement statistics

Values in parentheses are for the outer shell. ASU = asymmetric unit; R.m.s. = Root mean square.