Structural basis of diverse homophilic recognition by α- and β-protocadherins

Kerry Marie Goodman[†1,2], Rotem Rubinstein[†1,2,3], Chan Aye Thu[1], Fabiana Bahna[1,2,3], Seetha Mannepalli[1,2], Göran Ahlsén[2,3], Chelsea Rittenhouse[1], Tom Maniatis[1], Barry Honig[1,2,3,4*], and Lawrence Shapiro[1,2*]

[1]Department of Biochemistry and Molecular Biophysics, [2]Department of Systems Biology, [3]Howard Hughes Medical Institute, and [4]Department of Medicine, Columbia University, New York, NY 10032, USA.

[†] Equal contribution

[*] To whom correspondence should be addressed.

Summary

Homophilic interactions between clustered protocadherin protein isoforms (α-, β- and γ-Pcdhs) underlie self-recognition and non-self-discrimination between individual vertebrate neurons. How Pcdhs bind one another to mediate homophilic interactions, and how diverse Pcdh homophilic specificities are encoded, have not been determined. Here we report crystal structures for extracellular regions from four mouse Pcdh isoforms (α4, α7, β6 and β8), each revealing similar homophilic *trans* dimers, and each comprising primary intermolecular EC1:EC4 and EC2:EC3 interactions, with smaller distinct EC3:EC3 and EC2:EC2 interactions specific to α- and β-Pcdhs. Isoform-specific conservation of *trans* interface residues identified likely structural determinants of specificity, and mutation of such residues, along with their *trans*-interacting partner residues, created Pcdhs with new homophilic specificities. Together these data show how the sequences of Pcdhs define their diverse homophilic specificities and suggest a molecular mechanism for the evolution of Pcdh diversity.

Introduction

In functional neural circuits, axons and dendrites originating from the same neuron do not stably contact one another or their own cell body; however, they are free to interact with the cellular processes and surfaces of other neurons (Kramer and Kuwada, 1983; Kramer and Stent, 1985; Zipursky and Sanes, 2010; Zipursky and Grueber, 2013). This property of "self-avoidance", with permissive interactions among non-self neurons, is mediated by cell surface recognition mechanisms that generate high cell surface diversity such that each neuron acquires a distinct stochastically determined cell surface identity (Zipursky and Grueber, 2013; Zipursky and Sanes, 2010; Hattori et al., 2008; Matthews et al., 2007; Chen and Maniatis, 2013; Kostadinov and Sanes, 2015). Pioneering work in *Drosophila* has shown that the *Dscam1* gene, through stochastic alternative RNA splicing, encodes cell-surface protein isoforms with up to 19,008 distinct extracellular regions, each capable of highly specific homophilic recognition (Miura et al., 2013; Schmucker et al., 2000; Wojtowicz et al., 2007). Each neuron expresses a small but distinct stochastic repertoire of Dscam isoforms (Miura et al., 2013; Neves et al., 2004; Zhan et al., 2004). These Dscam1 recognition proteins confer sufficiently unique cell surface identities to mediate robust self-repulsion with permissive non-self engagement in the *Drosophila* nervous system (Hattori et al., 2007; Hattori et al., 2009). This mechanism appears to be common to insects and other arthropod invertebrates (Zipursky and Grueber, 2013).

In vertebrates, neuronal self-recognition and non-self discrimination is mediated, at least in part, by the clustered protocadherins (Pcdhs) (Chen and Maniatis, 2013; Zipursky and Grueber, 2013; Zipursky and Sanes, 2010; Lefebvre et al., 2012; Kostadinov and Sanes,

2015), a specialized family of cell-surface transmembrane cadherins (Suzuki, 1996). Like Dscam1 proteins, Pcdhs represent a family of similar but distinct protein isoforms that mediate strictly homophilic cell–cell recognition and are stochastically expressed to provide diverse single-neuron identities (Tasic et al., 2002; Wang et al., 2002; Yagi, 2013; Thu et al., 2014; Rubinstein et al., 2015). In humans and mouse, 52 and 58 Pcdhs, respectively, are encoded by the tandemly arranged *Pcdh*α, β, and γ gene clusters (Wu and Maniatis, 1999; Wu et al., 2001). The mouse *Pcdh* gene cluster encodes 14 α, 22 β, and 22 γ Pcdh protein isoforms. Each isoform is comprised of six extracellular cadherin domains (EC1–6), a transmembrane region, and, for α- and γ-Pcdhs, a short cytoplasmic extension (Wu and Maniatis, 1999). Single-cell specific expression of Pcdh isoforms is achieved through stochastic promoter choice (Tasic et al., 2002; Wang et al., 2002), as opposed to stochastic alternative splicing for invertebrate *Dscam1* (Schmucker et al., 2000). Each neuron is thought to express a small subset of Pcdh isoforms comprising a random repertoire of ~10 α, β, and γ isoforms, which are monoallelically expressed, along with constitutive biallelic expression of all 5 "C-type" isoforms, αC1–αC2 and γC3–γC5, which are divergent in sequence from other Pcdh isoforms (Esumi et al., 2005; Hirano et al., 2012; Kaneko et al., 2006).

Structure/function experiments from our laboratories have shown that Pcdhs emanate from the membrane surface as EC6-mediated *cis* dimers (Thu et al., 2014; Rubinstein et al., 2015). This observation is consistent with the requirement that α-Pcdhs form EC6-mediated *cis* complexes with β, γ or C-type Pcdh isoforms to enable their export to the plasma membrane surface (Thu et al., 2014). In contrast to the high specificity of *trans*

binding implied by their exclusively homophilic cell aggregation properties (Schreiner and Weiner, 2010; Thu et al., 2014), *cis* dimers are thought to form promiscuously between Pcdh isoforms (Schreiner and Weiner, 2010; Thu et al., 2014; Rubinstein et al., 2015). Our working model posits that Pcdh *cis* dimers form highly specific *trans* interactions with Pcdh dimers displayed on apposed cell surfaces (Rubinstein et al., 2015).

Although these studies provided important insights into the nature of Pcdh homophilic interactions, they did not reveal the detailed molecular basis of Pcdh *trans* binding and did not reveal how variations in the sequences of Pcdh isoforms define their homophilic binding specificity. While early domain-shuffling experiments for a subset of γ-Pcdhs suggested that the determinants of *trans* homophilic specificity localize to the EC2 and EC3 domains (Schreiner and Weiner, 2010), we subsequently demonstrated that *trans* homophilic interactions require domains EC1–4, with evidence for EC2:EC3 and EC1:EC4 interfaces, and an overall head-to-tail organization of the *trans* homodimer (Rubinstein et al., 2015). Mutation correlation studies evaluated in light of crystal structures of two EC1–EC3 Pcdh fragments also suggested an EC1–EC4 *trans* interface (Nicoludis et al., 2015).

Here we report crystal structures for α- and β-Pcdh extracellular regions, each engaged in homophilic recognition, revealing the atomic-level basis of highly specific *trans* recognition between Pcdhs from apposed cell surfaces. Structures of Pcdhs $\alpha4_{EC1–4}$, $\alpha7_{EC1–5}$, $\beta6_{EC1–4}$, and $\beta8_{EC1–4}$ each depict cognate homophilic *trans* dimer complexes.

Their conformations are highly similar, each binding in head-to-tail orientation as predicted (Rubinstein et al., 2015; Nicoludis et al., 2015), and comprise EC1:EC4, EC2:EC3, and small α-specific EC3:EC3 and distinct β-specific EC3:EC3 and EC2:EC2 interactions. The intermolecular interfaces for these Pcdh *trans* dimers are extensive, burying between 3904 and 5071 $Å^2$, revealing interaction surfaces mainly populated by residues that vary among alternate α- and β-Pcdh isoforms, alongside some that are constant among alternate α- and β-Pcdh isoforms. We further show that mutation of interacting variable interfacial residues, which code for specificity within each Pcdh subfamily, generates new specificities. Considering the conservation of the clustered Pcdhs, from cephalopod invertebrates to man (Albertin et al., 2015), and the critical role Pcdhs play in the development of neural circuits (Chen and Maniatis, 2013; Kostadinov and Sanes, 2015), an understanding of the nature of Pcdh homophilic specificity at the atomic level is of fundamental importance.

Results

Structure Determination

Using a HEK-293 suspension-cell protein expression system, we produced a series of Pcdh ectodomains, each encompassing the *trans* interaction region, which we previously showed to comprise domains EC1–4 (Rubinstein et al., 2015). We produced both EC1–4 (Pcdhs α4, β6, β8) and EC1–5 (α7) proteins, omitting the EC6 domain, which biophysical analysis previously showed to mediate *cis* dimer interactions between Pcdhs (Rubinstein et al., 2015), and thus complicated the behavior of recombinant proteins.

We confirmed that these two α- and two β-Pcdh ectodomain fragments formed dimers in solution by sedimentation-equilibrium analytical ultracentrifugation. The dimer dissociation constants for $\alpha4_{EC1–4}$ and $\alpha7_{EC1–5}$ were 37 ± 8.2 nM and 2.91 ± 0.55 μM respectively (Table S2 and Rubinstein *et al*, 2015, respectively). The β-Pcdhs showed weaker binding, with respective dissociation constants for $\beta6_{EC1–4}$ and $\beta8_{EC1–4}$ of 16.3 ± 2.1 μM and 24.0 ± 0.43 μM (Table S2).

We determined crystal structures for these Pcdh ectodomain fragments, each of which adopted dimeric conformations geometrically consistent with *trans* cell–cell recognition. Structures of $\alpha4_{EC1–4}$, $\alpha7_{EC1–5}$, $\beta6_{EC1–4}$ and $\beta8_{EC1–4}$ were determined to 2.8, 3.6, 2.9 and 2.9 Å resolution, respectively (Table S1 and Experimental Procedures). We first determined the crystal structure of $\alpha7_{EC1–5}$, for which phases were determined by single-isomorphous replacement with anomalous scattering (SIRAS) with a platinum-derivative crystal. All subsequent structures were determined by molecular replacement. Data collection and refinement statistics are shown in Table S1.

Overall structural features of protocadherin ectodomains

The four Pcdh ectodomain structures reported here are highly similar overall (Table S3). As expected, each comprises either four or five β-sandwich EC domains, arranged in tandem and connected by inter-domain linker regions (Figure 1), all of which bind three Ca^{2+} ions via the canonical cadherin Ca^{2+}-binding motif as described previously (Rubinstein et al., 2015; Nicoludis et al., 2015). The structures reported here are the first to contain EC4 and EC5, which adopt the canonical EC domain structure (Figure S1).

The overall arrangement of EC domains in each structure provides Pcdhs with straight, rod-like ectodomains; this structure is distinct from that of classical cadherins, which are curved such that the first and last EC domains are rotated relative to each other by ~90° (Boggon et al., 2002; Harrison et al., 2011). For Pcdhs, angles between neighboring EC domains deviate from co-linear between 6° and 22°, and deviations from linearity are arranged so as to mostly cancel out one another, producing an overall linear structure for EC1–4.

All four structures are decorated with N- and O-linked glycans on surface regions away from the *trans* dimer interfaces (Figure 1). Single O-mannose residues—a glycan that appears to be specific to cadherin-superfamily proteins (Vester-Christensen et al., 2013; Lommel et al., 2013)—were observed at positions 194 and 196 (α7 numbering) of EC2 in the α7, β6 and β8 structures. These glycans were previously identified by mass spectrometry in Pcdhs αC2, β1, γA8 and γC5, and S/T residues are conserved at these positions among most Pcdhs (Vester-Christensen et al., 2013; Rubinstein et al., 2015). Additionally, the β8 structure shows an additional O-mannose on the subsequent solvent exposed residue T199. The α-Pcdh structures also reveal O-mannosylation of EC4 and EC5 residues. α4 EC4 contains three O-mannoses on adjacent solvent exposed residues: T409, S411 and S413. T409 is also O-mannosylated in the α7 structure, as is the EC5 residue S449. The function of these O-mannose residues is currently unknown.

All structures show head-to-tail Pcdh *trans* recognition dimers

The structures of the *trans* dimers from the four different Pcdh isoforms are similar overall. They each reveal head-to-tail dimers, with partner molecules positioned as if emanating from adjacent cell surfaces (Figure 1). These head-to-tail interactions are similar overall to the two-fold symmetric four interaction-domain arrangement of receptor protein tyrosine phosphatase μ (Aricescu et al., 2007). Among cadherins and protocadherins one complex structure—between cadherin-23 and protocadherin-15—has an antiparallel orientation (Sotomayor et al.. 2012), but the interfaces comprise only EC1 and EC2, and appear unrelated to Pcdhs.

As described below, there are notable differences between the α- and β-Pcdh dimers sufficient to provide family-wide recognition specificity between alternate α- and β-Pcdhs. For the $\alpha4_{EC1–4}$ and $\alpha7_{EC1–5}$ dimer structures, the RMSD is 1.9 Å. The RMSDs amongst the three $\beta8_{EC1–4}$ dimers in the asymmetric unit are between 1.5−2.1 Å, a similar range to the RMSDs between the $\beta6_{EC1–4}$ and $\beta8_{EC1–4}$ dimer structures of 1.9−2.5 Å. In contrast, RMSDs between α- and β-Pcdh dimers are significantly larger, ranging between 3.6 and 4.8 Å (Figure S2). This is due primarily to a difference in the relative rotations of the dimer-mate protomers in the α- and β-Pcdh *trans* dimers. Superimposing one protomer from each of the dimers reveals a family-specific ~30° difference in orientation of the dimer-partner molecule between the α- and β-Pcdh dimers (Figure S2).

Each *trans* recognition dimer is composed of distinct sub-interfaces: large EC1:EC4 and EC2:EC3 interfaces common to both subfamilies, small EC3:EC3 interfaces distinct in α-

and β-Pcdhs, and a small EC2:EC2 interface found only in β-Pcdhs. Since the EC2:EC3 region abuts the dimer 2-fold axis, the two symmetry-related EC2–3 interface regions per complex form a nearly contiguous large (1975-2826 Å) interface centered on the symmetry axis. The EC1:EC4 interaction, in contrast, forms a separate interface region ~70 Å distant (center-to-center). Altogether, large areas of molecular surface are buried in the *trans* dimer for each Pcdh structure: 4619 $Å^2$ for $\alpha4_{EC1–4}$, 3904 $Å^2$ for $\alpha7_{EC1–5}$, 4762 $Å^2$ for $\beta6_{EC1–4}$ and 4893–5071 $Å^2$ for $\beta8_{EC1–4}$.

Conserved isoform-specific interface residues are likely specificity determinants

To help identify molecular regions important to establish specificities among α- and β-Pcdh isoforms, we produced sequence logos from multiple-sequence alignments of α- and β-Pcdh isoform-orthologs from numerous mammalian species (Extended Experimental Procedures). To focus on regions relevant to specificity, we extracted the residue positions located within the homodimer interfaces to produce sub-logos (Figures 2D, 2G, 3E, and 4D). These show that some interfacial residues are conserved for all isoforms; some residues vary among isoforms, but fail to show consensus among species; and some are both isoform-specific and also conserved among species. For example, interfacial residues 36 and 38 are QD in α8, but RA in other α-Pcdhs for >90% of vertebrate species (Figure 2D). Such isoform-specific species-conserved interfacial residues are highly likely to play a role in binding specificity, although those that are isoform-specific but not conserved among species could also play specificity roles in individual species. Below, we interpret the Pcdh recognition dimer crystal structures in

light of the sequence properties revealed by the logo analysis, and test the logo-identified putative specificity residues through a functional mutational analysis.

The "opened-up" depictions of the $\alpha7_{EC1–5}$ and $\beta8_{EC1–4}$ homodimer interfaces (Figure 1C and F), highlight the α- and β-Pcdh constant and variable interfacial residues from the logo analysis. This reveals that variable residues—likely specificity residues—generally make intermolecular contacts with other variable residues on the partner molecule, while α or β constant residues generally contact other constant residues. Constant residues contribute ~50% of the buried surface area for α-Pcdhs, but only 5% of the buried surface area in the β-Pcdh dimers (Table S4). Part of this remarkable difference arises from the near identical sequences of α-Pcdh EC1:EC4 interface residues.

α-Pcdh EC1:EC4 interface

In the α-Pcdh structures, the EC1:EC4 interface (Figures 2A–D) buries ~975 $Å^2$, thereby contributing ~1950 $Å^2$ buried surface area overall via the two copies of EC1:EC4 present in each of the α-Pcdh dimers (Table S4). The EC1:EC4 interface of α4 and α7 are shown in Figure 2B (top and bottom, respectively). The α-isoform logo alignment in Figure 2D identifies interface residues as either constant or variable among alternate vertebrate α-Pcdhs; the interface is shown "opened out" and displayed as a surface in Figure 2C, with constant residues shown in orange and variable residues in purple.

The EC1:EC4 interface runs over about half the length of the cadherin domains, involving the N-terminal end of the EC1 C, F, and G-strands and the C-terminal end of

the A, D, E, and B-strands of EC4. Constant interactions include a cluster of conserved aliphatic interface residues from the EC1 FG region (I80, P84, L85, and V87) and the B and D-strands of EC4 (T338, V339, L342, and T371) (Figure 2B and C, left side). Toward the C-terminal end of EC1 (Figures 2B and C, right side) constant elements of the interface involve numerous polar and charged residues, including EC1 residues K40 (CD-loop), H76 and E78 (F-strand), and H89 and E91 (G-strand), which form conserved interactions with EC4 residues F372, K373, and Y375 (DE-loop). Overall, these constant:constant interactions account for ~70–88% of the EC1:EC4 interface buried surface area (Table S4).

Elements of the EC1:EC4 interface that vary between α-isoforms are localized to the edge of the interface, where variable residues from the C-strand of EC1 pack against variable residues of the EC4 A-strand. Additionally the α4 structure reveals two isoform-specific salt bridges between constant and variable residues ($E78_{constant}$:$K324_{\alpha3,4,8}$ and $K40_{constant}$:$E322_{\alpha4}$) (Figure 2D).

β-Pcdh EC1:EC4 interface

The β-Pcdh EC1:EC4 interface involves similar residues and structural elements to the α-Pcdh EC1:EC4 interface, although there is a greater contribution to the interface from the EC1 C-strand and CD-loop, and no contribution from the EC1 A-strand (Figure 2E). The β-Pcdhs have a slightly larger EC1:EC4 interface than the α-Pcdhs, contributing 2260 and 2120 $Å^2$ buried surface area respectively to the β6 and β8 *trans* dimers. In contrast to α-Pcdhs, almost all of the interface residues vary among β isoforms; a number of these

are conserved in the orthologous isoforms of other vertebrate species, underscoring their likely function in recognition specificity (Figure 2F and 2G).

β6 and β8 are the most closely related β-Pcdhs with 92% sequence identity in their EC1–4 regions. Consequently their EC1:EC4 interfaces (Figure 2E) are nearly identical, with only three differences in the identity of the interfacial residues, two of which interact R/N41:E/K369. One further difference in the β6 and β8 EC1:EC4 interfaces is the relative positions of EC4 A-strand residues S322 and L323, which differ despite their conserved identities in β6 and β8. This positional difference enables the S322 side chain to partake in the dimer interface of β6, but not β8, and conversely the L323 side chain makes dimer contacts in β8, but not β6.

α-Pcdh EC2:EC3 interface

The EC2:EC3 interface is highly variable for all three Pcdh families and appears to represent the primary specificity region for α-Pcdhs, with variable residues accounting for 83–88% of the EC2:EC3 buried surface area. Taken as a whole, the symmetric EC2–3:EC2–3 interface region buries 2627 $Å^2$ in the α4 structure and 1975 $Å^2$ in the α7 structure, involving contacts between the ABED-face of EC2 and the AGFC-face of EC3 (Figure 3B).

The majority of EC2:EC3 interfacial area in both α-Pcdh structures is provided by interaction of the EC2 AB-loop and start of the EC2–3 linker region, with the EC3 FG-loop and the C-terminal end of the EC2–3 linker region (colored pale purple and orange

in Figure 3D). Aside from the conserved calcium-coordinating linker residues (D205, D208 and N209), over half of the surface residues in these contacting regions differ between α4 and α7 (Figure 3B and 3E). In both structures, the majority of EC2-AB/EC3-FG-loop residues in contact are hydrophobic, with the only charged residue that does not coordinate calcium, EC2 AB-loop residue R117, forming a hydrogen bond with EC3 FG-loop residue S302 in α4 and with Q300 in α7. This region also contains one notable α-Pcdh constant interaction between H304 (EC3 G-strand), and EC2 B-strand residues F123 and P124, which is conserved in all isoforms except α3, where it is serine (Figure 2D–E).

There are two additional interaction regions that make up the EC2:EC3 interface. Firstly there is a small EC2 A-strand:EC3 CD-loop contact, which, in both α4 and α7, is a salt bridge formed between K109 (EC2) and D255 (EC3). This interaction is conserved across vertebrate species in α2, α3, α8 and α10, but not the other α-Pcdhs. Secondly there is an EC2 DE-loop:EC3 AGF-sheet interaction, which is the most divergent region of the EC2:EC3 interface since, as well as being highly variable in sequence (Figure 2E), the EC2 DE-loop also adopts different conformations in the α4 and α7 structures, resulting in differing interactions with the EC3 AGF-sheet surface (Figure 2B). Poor electron density for the DE-loop regions in both structures suggest that the EC2 DE-loop is more mobile than other α-Pcdh interfacial elements. Interestingly, a number of the contacts in this interface are main-chain mediated. The role this region could play in α-Pcdh specificity is therefore unclear.

β-Pcdh EC2:EC3 interface

The β-Pcdh EC2:EC3 interface involves the EC2 ABED-face and EC3 GFC-face, similar to α-Pcdhs. Nonetheless, the β-Pcdh EC2:EC3 interface has marked differences, and involves both residues in common with the α-interface as well as residues that are distinct (Figure 4B). In common with α-Pcdhs, the β-Pcdh EC2:EC3 interface is highly diversified amongst β isoforms, with 92–93% of the buried surface area contributed by variable residues (Figure 4D and 4E and Table S4). However the, β6 and β8 isoforms differ only in the identity of two EC2:EC3 interface residues, both of which are in EC2 (S/I117 and L/P125). The EC2:EC3 interface buries 2504 $Å^2$ in the β6 dimer and 2826 $Å^2$ in the β8 dimer.

The β-Pcdh EC2 AB-loop and EC3 FG-loop form interactions similar to those observed in the α structures; although the contacts are not as extensive in β-Pcdhs because the EC3 FG-loop is two residues shorter and the apex of the loop consists of two glycine residues in β6 and β8 (three in most other β isoforms). This allows Q115, near the apex of the EC2 AB-loop to form hydrogen bonds with the EC2–3-linker region, rather than specific contacts with the FG-loop as the equivalent residue (114) does in the α structures (Figure 4B). The contacts between the EC2 B-strand residues 122–125 and the EC3 F- and G-strands are also similar to those of the α-Pcdhs. However, unlike α-Pcdhs, the CD-loop makes extensive contacts in the interface, both with the remainder of the EC2 B-strand (residues 127–132) and the EC2 DE-loop (Figure 4B). The CD-loop contains four glycine residues in β6 and β8, but not in any other β-Pcdh (Figure 4E), and therefore may

not be able to insert so readily between the EC2 B-strand and the DE-loop in other β-Pcdh isoforms. The EC2 DE-loop also engages EC3 F- and G-strand residues, like in the α's, although the DE-loop is not so closely packed to the EC3 surface resulting in more side chain:side chain interactions rather than main chain:side chain as seen in the α structures.

Small EC2:EC2 and EC3:EC3 interfaces at the *trans*-dimer 2-fold

For the α-Pcdhs, a small region of EC3:EC3 intermolecular interaction is formed by symmetry-related contacts between the side chains of FG-loop residue 298 from each protomer of the *trans* dimer (Figure 3C). In α7 P299 also contributes to this interaction alongside the hydrophobic F298:F298 interaction. Notably the EC3 FG-loop is two residues shorter in β-Pcdhs, too short to make contact at the 2-fold axis.

The β-Pcdh ectodomain structures indicate that small interaction regions at the *trans* dimer two-fold axis—involving the self-interactions of the EC3 BC-loop residue 246 and the EC2 AB-loop residue 117—will occur when these residues are large, as is the case for some but not all β-Pcdhs (Figure 4C). Residue 246 does not make self-contacts in the β6 and β8 structures, where it is a histidine in both. However 246 is a phenylalanine or tyrosine in 12 of the 22 mouse β-Pcdhs, which are both large enough that this residue would be expected to self-interact in the *trans* dimer, and therefore contribute to the binding energy and potentially to specificity.

Residue I117 self-interacts in the β8 crystal structure, but S117 in the β6 structure does not since the serine is not large enough to span the gap between protomers. Residue 117 is also an isoleucine in β13, and an asparagine, which could also self-interact, in 10 other mouse β-Pcdhs. As mentioned above, the only differences in EC2:EC3 interfacial residues between β6 and β8 occur in EC2 and one of these differences is S/I117. This is consistent with previous data showing that a β6/β8 chimera containing EC1–2 from β8 and EC3–6 from β6 was able to intermix freely with β8. This differs from the usual requirement for three out of the four interacting domains to be matched to convert interaction specificity (Schreiner and Weiner, 2010; Thu et al., 2014; Rubinstein et al., 2015). The basis for this EC1–2-only specificity conversion can be explained, at least in part, by the self-interaction of residue 117.

Mutants of *trans*-interacting residues produce new protocadherins

In order to identify residues important for recognition specificity, we performed mutagenesis experiments both in cells and *in silico*. We used logo analysis to identify interface residues that are conserved across species in individual Pcdh isoforms, then evaluated residue interactions across the *trans* interface, and chose pairs or clusters of interacting residues that had significant differences in chemical character in different isoforms. We then mutated these pairs or clusters of complementary *trans*-interacting residues, in a full-length Pcdh-α7 or β6 context, to those from a different α or β isoform and assessed whether a new specificity was generated using the K562 cell aggregation assay (Thu et al., 2014). Cells transfected with these mutants were mixed with cells expressing either wild-type α7 or β6. We then assessed whether the mutant and wild-type

cell populations formed separate aggregates – which would indicate that the mutant Pcdh preferentially recognized itself – or alternatively formed mixed aggregates – which would indicate that the mutant did not create a new Pcdh specificity (Figure 5A). In addition, we produced models of heterophilic *trans* dimers for each mutant-Pcdh bound to the wild-type Pcdh, as well as the homophilic complexes formed by each mutant, to assess the likely structural impact of the mutation.

For example, in the case of α-Pcdhs, and focusing on the EC1:EC4 interaction, the logos show that only two residues in EC1 have conserved changes between isoforms (Figure 2D). We chose EC1 residue 36, which is a glutamine in α8 and an arginine in all other α-Pcdhs. In α7, R36 interacts with the EC4 residue S327, and therefore in α8, Q36 would be expected to interact with W327, which is also exclusively present in α8 across species. We generated the α7 R36Q/S327W double mutant, and assessed its recognition properties. α7 R36Q/S327W expressing cells formed mixed aggregates with wild-type α7 expressing cells, indicating that these mutations were insufficient to generate a new specificity (Figure 5Bi). We therefore expanded the area around these residues to produce a new mutant in which residues $36–38_{EC1}$ and $322–327_{EC4}$ from α8 were swapped into α7. In this case, mutant and wild-type expressing cells formed separate aggregates, indicating that the mutant and wild-type Pcdhs displayed preferential homophilic specificity (Figure 5Bii). Modeling the homophilic interface of the mutant α7 protein containing these nine α8 interface residues, suggested the favorable formation of two salt bridges involving the α8-specific residue K324. Modeling the heterophilic mutant:wild-type interface showed an unfavorable interaction between R36 and W327. Together,

these favorable homophilic and unfavorable heterophilic interactions appear to explain the change in specificity (Figure 5Biii). Importantly, mutating the EC4 residues 322–327 alone, without mutating their counterpart interface residues in EC1, resulted in mixed aggregates of mutant and wild-type α7 expressing cells, indicating that single-sided mutations were insufficient to generate a new specificity.

To probe the α-Pcdh EC2:EC3 interface, we first chose the interacting residues 109_{EC2} and 255_{EC3}. These residues show correlated changes in the logo analysis (Figure 3E), where the K_{109}:D_{255} pair is conserved in α2, α3, α4, α7, α8, and α10, while Q_{109}:L_{255} is conserved in α5, α6, and α9. We generated both an α7 single mutant (D255L) and an α7 double mutant of the interacting pair (K109Q/D255L), and tested their binding specificities with the parent α7 isoform. Cells expressing the α7 single mutant formed mixed aggregates with cells expressing wild-type α7, while cells expressing the double mutant formed separate aggregates from those expressing wild-type α7 (Figure 5Ci-ii). Consistent with these observations, the model of the homophilic double mutant complex showed no clashes, while the heterophilic complex model with wild-type α7 indicated that a new specificity was likely formed due to electrostatic incompatibility between residues L255 and K109 resulting in unfavorable heterophilic interactions (Figure 5Ciii).

Similarly, we introduced mutations in the interacting residues 114_{EC2} and 301_{EC3} from the α-Pcdh EC2:EC3 interface, which we changed to the corresponding residues in α3 (Figure 5D). Mutating these residues did not result in a new specificity (Figure 5Di). Based on the structure, we mutated an expanded area around residue 301 so as to produce

an additional mutant in which residues 300 and 302 were also changed to α3 residues. This mutant generated a new Pcdh specificity (Figure 5Dii). Modeling of the homophilic complex for this mutant isoform suggested that the homophilic mutant:mutant interface had compatible van der Waals interactions, as well as a favorable hydrogen bond between R117 and T302. In contrast, the heterophilic interface with wild-type α7 lacked this hydrogen bond, and had incompatible van der Waals interactions between Y114 and Q300 (Figure 5Diii). We also tested self-interacting α7 EC3:EC3 interface residue 299 (Figure 5Ei). Mutating only this single residue (P299F) generated a new specificity (Figure 5Ei). Modeling confirmed the possibility of F299 self-interacting (Figure 5Eii).

We next carried out a similar mutational analysis for the β-Pcdh interfaces. Mutations of *trans*-interacting residues 39_{EC1} and 41_{EC1}, with 340_{EC4} in the EC1:EC4 interface, were sufficient to create a new Pcdh specificity (Figure 6Ai). In addition, three experiments involving exchanges of interacting residue pairs in the β6 EC2:EC3 interface with residues from β19, β10, or β7, respectively, were each sufficient to produce new Pcdh specificities (Figure 6B-D). Modeling suggested that the homophilic interactions for the EC1:EC4 interface mutant H39V/R41D/S340R, as well as two of the EC2:EC3 interface mutants (K113E/S295R and R157N/E289K), were favored by electrostatic complementarity between the interacting mutated residues. In contrast, interactions of all three of these mutants with wild-type β6 would position residues with the same charge in close proximity (Figure 6 Aii, Bii, and Dii). Modeling of the final EC2:EC3 interface mutant, Q115P/L298F, suggested that the new specificity observed in cell aggregation experiments resulted from van der Waals clashes, making heterophilic interaction with

wild-type β6 unfavorable (Figure 6 Cii). In addition, we tested the self-interacting β-Pcdh EC2:EC2 interface residue 117 near the two-fold axis of the dimer interface. The β6 S117I mutant resulted in formation of separate cell aggregates, indicating that this single mutation was sufficient to generate a new specificity (Figure 6E).

Discussion

We showed previously that α, β, and γ-Pcdhs, when expressed in non-adherent cells and appropriately transported to the cell surface, mediate cell aggregation with strictly homophilic specificity (Thu et al. 2014). The crystal structures reported here reveal the atomic-level basis of this specific self-binding, revealing canonical modes of *trans* interaction that are similar but distinct for α- and β-Pcdhs. *Trans* recognition is mediated primarily by large EC2:EC3 and EC1:EC4 interactions, as well as small EC3:EC3 interfaces for α-Pcdhs and both EC2:EC2 and EC3:EC3 interfaces for β-Pcdhs. The structures explain the results of prior mutagenesis studies (Schreiner and Weiner, 2010; Thu et al., 2014; Rubinstein et al., 2015) and mutation correlation analyses (Nicoludis et al., 2015). Moreover, new structure-guided mutants reported here show that *trans* interface residues that vary among isoforms, but are conserved in a given Pcdh isoform across multiple species, play important roles in homophilic specificity. Together with our previous mutagenesis results (Rubinstein et al., 2015), these data show that the combined specificity of the EC1:EC4 and EC2:EC3 interfaces, the EC2:EC2 interface in β-Pcdhs, and the EC3:EC3interface in α-Pcdhs, underlie the homophilic recognition specificities of Pcdh isoforms.

The Pcdh *trans*-binding mode is completely different from that of the well-characterized classical cadherins. For classical cadherins, a 90° curve in the ectodomain enables a parallel *trans* binding configuration with domain:domain self-binding. Classical cadherin recognition interfaces are contained within EC1 burying comparably small surface areas (~1600–1800 $Å^2$ for type I and ~2700–3300 $Å^2$ for type II cadherins). They are formed by a core of conserved residues surrounded by specificity residues, which account for ~50% of interfacial surface area for the five type I cadherins (Boggon et al., 2002; Harrison et al., 2011; Patel et al., 2006; Vendome et al., 2014). Pcdh ectodomains, in contrast, are remarkably straight; they bind in *trans* in an antiparallel head-to-tail orientation such that each adhesion dimer includes two symmetry-related copies each of the EC1:EC4 and EC2:EC3 interfaces, which bury up to 5071 $Å^2$ in total. Up to 96% of the interfacial surface area in Pcdh recognition interfaces involves residues that vary between isoforms to define binding specificity (Table S4). Thus, compared with classical cadherins, Pcdhs have larger and more varied *trans* interfaces, consistent with the higher number of Pcdh isoforms that must be distinguished.

The logo analyses shown in Figures 2, 3 and 4 reveal many binding-interface residues that vary among isoforms, but are evolutionarily conserved for each isoform, suggesting their likely importance in defining binding specificity. We used this observation as the basis of functional studies in which we mutated these conserved isoform-specific residues, along with their *trans*-interacting partner residues – most of which also showed isoform-specific conservation – to determine if mutation of the interacting pairs were sufficient to change Pcdh binding specificity. While some such mutants (e.g. Figure 5B)

yielded wild-type and mutant transfected cell populations that formed mixed aggregates – indicating they were insufficient to change binding specificity – others led to the formation of separate cell aggregates, indicating a new Pcdh specificity had been created. Thus, interface residues identified by their isoform-specific conservation effectively formed "new" protocadherins that exhibited both homophilic specificity and the ability to distinguish between self (the mutant) and non-self (the parental Pcdh). A recent study showed that the bacterial ParD-ParE toxin-antitoxin system evolved specific protein-protein interactions through promiscuous intermediate proteins, rather than through non-interacting intermediate proteins that were rescued by compensatory mutations (Aakre et al., 2015). Interestingly, the mutagenesis results presented here, as well as results from domain swap mutagenesis in our previous study (Rubinstein et al., 2015), show that mutating only one side of the Pcdh interface, e.g. EC1 without its *trans*-interacting partner residues in EC4, typically resulted in a functional isoform with promiscuous specificity, able to recognize both self and parent wild-type isoforms. This is likely to at least partly mirror the process by which Pcdhs diversified in evolution, where initial mutations could result in promiscuous interactions, and a second interacting mutation could then generate a novel highly specific homophilic interaction.

While the structures of Pcdhs are different from those of Dscam1 isoforms, which mediate homophilic recognition underlying neuronal self-avoidance in insects and other arthropods, there appear to be commonalities in the character of Pcdh and Dscam1 domain:domain interactions. Recognition between Dscam1 proteins is mediated by three self-to-self Ig-domain interface regions, Ig2:Ig2, Ig3:Ig3, and Ig7:Ig7, which are

diversified by alternative splicing (Sawaya et al., 2008). There are 12 different Ig2 isoforms, 48 different Ig3 isoforms and 33 different Ig7 isoforms. Thus, at the structural level, each of the three interfaces must encode sufficient variability to distinguish 12, 48 and 33 isoforms for Ig2, Ig3 and Ig7, respectively using ~850–1330 $Å^2$ buried surface area per domain. Ig3 has the most difficult problem to solve, as it must distinguish among the largest number of isoforms using only 850 $Å^2$ of buried surface area. Although Pcdh:Pcdh interfaces bury approximately 4000–5000 $Å^2$, only half of the two-fold symmetric interface is available for the generation of diversity since the EC1:EC4 and EC2:EC3 interfaces are duplicated. Pcdhs thus succeed in discriminating among 58 isoforms using ~2000–2500 $Å^2$ of unique buried surface area. Variable residues in Pcdhs and Dscam1 isoforms account for similar percentages of interfacial surface area (~80% for Dscam1 and 47–96% for Pcdhs). Thus, in terms of the diversity encoded per Angström interfacial surface area, Pcdhs are similar to Dscams (58×58 over 2000–2500 $Å^2$ versus 12×12, 48×48 and 33×33 over 850–1330 $Å^2$ each).

Recognition between Dscam1 isoforms depends on precisely matched recognition domains (Zipursky and Grueber, 2013). A truncated Dscam1 molecule containing only the N-terminal horseshoe structure that includes the Ig2:Ig2 and Ig3:Ig3 interfaces was monomeric (Hughes et al., 2007). Additionally, high-throughput binding experiments with Dscam1 protein isoforms showed that matches in all three specificity domains were required for binding in the vast majority of cases (Wojtowicz et al., 2007). For Pcdhs, EC1–3 fragments do not dimerize despite the fact that they contain the regions necessary for the EC2:EC3 recognition interface to form (Rubinstein et al, 2015; Nicoludis et al.,

2015). Additionally, experiments with Pcdh chimeras showed that matches in both EC1:EC4 and EC2:EC3 interfaces were required to mediate cell aggregation in transfected-cell assays (Rubinstein et al., 2015). Thus for Pcdhs, as for Dscam1 isoforms, homophilic specificity derives from the combination of all variable interface regions.

Stochastic expression of thousands of Dscam1 isoforms is sufficient to solve the problem of self-recognition and non-self discrimination in arthropod invertebrate nervous systems (Hattori 2009). In contrast, only ~60 stochastically expressed Pcdh isoforms (e.g. in mouse and human) solve a similar problem at larger scale in vertebrates. The combination of promiscuous EC6-mediated *cis* and specific *trans* interactions (Thu et al., 2014) led us to propose formation a *trans*-cellular zipper-like assembly of Pcdhs whose growth could be terminated by isoform mismatches (Rubinstein et al., 2015). In principle, expansion of recognition diversity could be accomplished through this mechanism. Such an assembly could potentially bring Pcdh cytoplasmic regions, and their associated proteins, into apposition to activate downstream signaling pathways. This signal could only be generated in the presence of an apposed neuronal surface with an identical Pcdh isoform repertoire—the natural representation of "self".

As demonstrated by the structures reported here, the α- and β-Pcdhs have similar interface regions, with specificity elements localized to corresponding structural regions. Similarities in sequence combined with similar effects of domain deletions, swaps, and point mutations, is suggestive of a related mechanism for γ-Pcdhs, and potentially C-type

Pcdhs as well. It remains possible, however, that specific assemblies dependent on the properties of each subfamily could be required for the function of the clustered Pcdhs.

Experimental Procedures

Protein production and crystallography

Pcdh protein fragments $\alpha4_{EC1–4}$, $\alpha7_{EC1–5}$, $\beta6_{EC1–4}$ and $\beta8_{EC1–4}$ were expressed in suspension HEK293 cells (Invitrogen). Crystals were grown by vapor diffusion in: 3 M sodium chloride, 0.1 M Mes pH 6.2 for $\alpha4_{EC1–4}$; 14% PEG3350, 0.1 M Hepes pH 7.5 for $\alpha7_{EC1–5}$; 23% PEG3350, 0.2 M potassium thiocyanate for $\beta6_{EC1–4}$; and 15.3% PEG550MME, 7.7% PEG20000, 30 mM calcium chloride, 30 mM magnesium chloride, 0.1 M Tris/Bicine buffer pH 8.5 (Molecular Dimensions) for $\beta8_{EC1–4}$. The $\alpha7_{EC1–5}$ crystals were dehydrated against 35% PEG3350, 0.1 M Hepes pH 7.5 to improve stability. The $\alpha7_{EC1–5}$ crystal structure was solved using the SIRAS technique with a platinum-derivative crystal, and the other structures were solved by molecular replacement. See Extended Experimental Procedures for details.

Co-aggregation assay to test *trans* binding specificity.

Cell aggregation assays were performed as previously described in Thu et al., 2014. mCherry- or mVenus-tagged wild type or mutant Pcdh expression constructs were transfected into K562 cells (human leukemia cell line, ATCC CCL243). Transfected cells were mixed after 24 h by shaking for 1–3 h. Cell aggregates were then imaged. See Extended Experimental Procedures for details.

Structural modeling

Models of homophilic complexes formed by both the wild-type and mutant isoforms, as well the heterophilic dimer between mutant and wild-type isoforms were generated using the crystal structures of Pcdhα7, Pcdhα4 and Pcdhβ8 as structural templates. Side chains of mutated residues were first modified to alanine and then modeled using Scwrl (Krivov et al., 2009). Importantly, the three models were generated using the same procedure and therefore, there was no bias towards homophilic or heterophilic models.

Accession numbers

Atomic coordinates and structure factors have been deposited to the protein data bank with accession codes 5DZW, 5DZV, 5DZX and 5DZY for $\alpha4_{EC1–4}$, $\alpha7_{EC1–5}$, $\beta6_{EC1–4}$ and $\beta8_{EC1–4}$ respectively.

Author contributions

K.M.G., R.R., C.A.T., T.M., B.H. and L.S. designed experiments, analyzed data and wrote the paper. F.B., S.M. and K.M.G. cloned, expressed, purified and crystallized all proteins. K.M.G. determined the crystal structures. S.M. and F.B. performed the site-directed mutagenesis. R.R. conducted the sequence and structural analysis of specificity. 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 Igor Kourinov and Surajit Banerjee for help with synchrotron data collection at the APS NE-CAT 24-ID-C/E beamlines, supported by NIH P41 GM103403. We also thank Chaim Schramm for help with preparing the sequence logo analysis. 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).

References

Aakre, C.D., Herrou, J., Phung, T.N., Perchuk, B.S., Crosson, S., and Laub, M.T. (2015). Evolving New Protein-Protein Interaction Specificity through Promiscuous Intermediates. Cell *163*, 594–606

Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., and Rokhsar, D.S. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature *524*, 220–224.

Aricescu, A.R., Siebold, C., Choudhuri, K., Chang, V.T., Lu, W., Davis, S.J., van der Merwe, P.A., Jones, E.Y. Structure of a tyrosine phosphatase adhesive interaction reveals a spacer-clamp mechanism. Science *317*,1217–1220.

Boggon, T.J., Murray, J., Chappuis-Flament, S., Wong, E., Gumbiner, B.M., and Shapiro, L. (2002). C-cadherin ectodomain structure and implications for cell adhesion mechanisms. Science *296*, 1308–1313.

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.*, et al.* (2012). Functional significance of isoform diversification in the protocadherin gamma gene cluster. Neuron *75*, 402–9.

Chen, W.V., and Maniatis, T. (2013). Clustered protocadherins. Development *140*, 3297–3302.

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.

Han, M.H., Lin, C., Meng, S., and Wang, X. (2010). Proteomics analysis reveals overlapping functions of clustered protocadherins. Mol. Cell Proteomics *9*, 71–83.

Harrison, O.J., Jin, X., Hong, S., Bahna, F., Ahlsen, G., Brasch, J., Wu, Y., Vendome, J., Felsovalyi, K., Hampton, C.M.*, et al.* (2011). The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins. Structure *19*, 244–256.

Hattori, D., Demir, E., Kim, H.W., Viragh, E., and Zipursky, S.L., and Dickson, B.J. (2007). Dscam diversity is essential for neuronal wiring and self-recognition. Nature *449*, 223–227.

Hattori, D., Millard, S.S., Wojtowicz, W.M., and Zipursky, S.L. (2008). Dscam-mediated cell recognition regulates neural circuit formation. Annu. Rev. Cell Dev. Biol. *24*, 597–620.

Hattori, D., Chen, Y., Matthews, B.J., Salwinski, L., Sabatti, C., Grueber, W.B., and Zipursky, S.L. (2009). Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. Nature *461*, 644–648.

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.

Hughes, M.E., Bortnick, R., Tsubouchi, A., Baumer, P., Kondo, M., Uemura, T., and Schmucker, D. (2007). Homophilic Dscam interactions control complex dendrite morphogenesis. Neuron *54*, 417–427.

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*, 30551–30560.

Kostadinov, D., and Sanes, J.R. (2015) Protocadherin-dependent dendritic self-avoidance regulates neural connectivity and circuit function. Elife *4*, e08964.

Kramer, A.P., and Kuwada, J.Y. (1983). Formation of the receptive fields of leech mechanosensory neurons during embryonic development. J. Neurosci. *3*, 2474–2486.

Kramer, A.P., and Stent, G.S. (1985). Developmental arborization of sensory neurons in the leech Haementeria ghilianii. II. Experimentally induced variations in the branching pattern. J. Neurosci. *5*, 768–775.

Krivov, G.G., Shapovalov, M.V., and Dunbrack Jr., R.L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins *77*, 778–795.

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.

Lommel, M., Winterhalter, P.R., Willer, T., Dahlhoff, M., Schneider, M.R., Bartels, M.F., Renner-Muller, I., Ruppert, T., Wolf, E., and Strahl, S. (2013). Protein O-mannosylation is crucial for E- cadherin-mediated cell adhesion. Proc. Natl. Acad. Sci. USA *110*, 21024–21029.

Matthews, B.J., Kim, M.E., Flanagan, J.J., Hattori, D., Clemens, J.C., Zipursky, S.L., and Grueber, W.B. (2007). Dendrite self-avoidance is controlled by Dscam. Cell *129*, 593–604.

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.

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 doi:10.1016/j.str.2015.09.005

Patel, S.D., Ciatto, C., Chen, C.P., Bahna, F., Rajebhosale, M., Arkus, N., Schieren, I., Jessell, T.M., Honig, B., Price, S.R., and Shapiro, L. (2006) Type II cadherin ectodomain structures: implications for classical cadherin specificity. Cell *124*, 1255–68.

Ribich, S., Tasic, B., and Maniatis, T. (2006). Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. Proc. Natl. Acad. Sci. USA *103*, 19719–19724.

Rubinstein, R., Thu, C.A., Goodman, K.M., Wolcott, H.N., Bahna, F., Mannepalli, 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.

Sawaya, M.R., Wojtowicz, W.M., Andre, I., Qian, B., Wu, W., Baker, D., Eisenberg, D., and Zipursky, S.L. (2008). A double S shape provides the structural basis for the extraordinary binding specificity of Dscam isoforms. Cell *19*, 1007–18

Schmucker, D., and Chen, B. (2009). Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. Genes Dev. *23*, 147–156.

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.

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.

Soba, P., Zhu, S., Emoto, K., Younger, S., Yang, S.J., Yu, H.H., Lee, T., Jan, L.Y., and Jan, Y.N. (2007). Drosophila sensory neurons require Dscam for dendritic self-avoidance and proper dendritic field organization. Neuron *54*, 403–416.

Sotomayor, M., Weihofen, W.A., Gaudet, R., and Corey, D.P. Structure of a force-conveying cadherin bond essential for inner-ear mechanotransduction. Nature *492*, 128–132

Suzuki, S.T. (1996). Protocadherins and diversity of the cadherin superfamily. J. Cell Sci. *109*, 2609–2611.

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.

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.

Vendome, J., Felsovalyi, K., Song, H., Yang, Z., Jin, X., Brasch, J., Harrison, O.J., Ahlsen, G., Bahna, F., Kaczynska, A., Katsamba, P.S., Edmond, D., Hubbell, W.L., Shapiro, L., and Honig, B. (2014). Structural and energetic determinants of adhesive binding specificity in type I cadherins. Proc. Natl. Acad. Sci. USA *111*, E4175–84

Vester-Christensen, M.B., Halim, A., Joshi, H.J., Steentoft, C., Bennett, E.P., Levery, S.B., Vakhrushev, S.Y., and Clausen, H. (2013). Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. Proc. Natl. Acad. Sci. USA *110*, 21018–21023.

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.

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.

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.

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.

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.

Wu, W., Ahlsen, G., Baker, D., Shapiro, L., and Zipursky, S.L. (2012). Complementary chimeric isoforms reveal Dscam1 binding specificity in vivo. Neuron. *74*, 261-268.

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*, 510–513.

Yagi, T. (2012). Molecular codes for neuronal individuality and cell assembly in the brain. Front. Mol. Neurosci. *5*, 45.

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

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.

Zipursky, S.L., and Grueber, W.B. (2013). The molecular basis of self-avoidance. Ann. Rev. Neurosci. *36*, 547–568.

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

Figure Legends

Figure 1. Crystal structures of the α- and β-Pcdh cell–cell recognition dimers

A. Crystal structure of the $\alpha4_{EC1–4}$ dimer. The two EC1–4 protomers (colored cyan and grey) bind in a symmetrical anti-parallel configuration with EC1 interacting with EC4 and EC2 interacting with EC3. Bound calcium ions are shown as green spheres and glycans are shown as red, blue and white spheres.

B. The $\alpha7_{EC1–5}$ structure, protomers colored green and grey, shows a near identical EC1–4 mediated head-to-tail dimer to α4 (RMSD = 1.9 Å). The EC5 domains extend laterally, and are therefore not involved in the dimer interaction.

C. Surface representation of the two $\alpha7_{EC1–5}$ protomers, opened out to expose the dimer interface, which is colored orange for residues that are constant among α-Pcdhs and purple for residues that vary between α-Pcdhs.

D. The $\beta6_{EC1–4}$ dimer, protomers colored yellow and grey, showing a similar EC1–4 mediated head-to-tail dimer to the α-Pcdh structures (RMSD to $\alpha4_{EC1–4}$ dimer = 4.8 Å).

E. The $\beta8_{EC1–4}$ dimer (chains A and B from the crystal structure are shown), protomers colored pink and grey, is near identical to the $\beta6_{EC1–4}$ dimer (RMSD = 1.6 Å)

F. Surface representation of the two $\beta8_{EC1–4}$ protomers, opened out to expose the dimer interface, which is colored orange for residues that are constant among β-Pcdhs and purple for residues that vary between β-Pcdhs.

See also Figures S1–S2 and Tables S1–S4.

Figure 2. The EC1:EC4 interface is largely conserved amongst α-Pcdhs and highly diverse amongst β-Pcdhs

A. The $\alpha7_{EC1–5}$ dimer structure with the EC1:EC4 interface highlighted (red box).

B. Close up of the EC1:EC4 interface in the $\alpha4_{EC1–4}$ (top panel, EC1 from one protomer is colored cyan the EC4 from the other protomer grey) and $\alpha7_{EC1–5}$ (bottom panel, EC1 green, EC4 grey) structures. Side chains are shown and labeled for all residues where the side chain contributes to the dimer interface.

C. Surface representation of α7 EC1 and EC4 interfacial regions from an opened out dimer, highlighting the putative specificity-determining residues. Interfacial residues are colored orange if they are conserved amongst all α-Pcdhs and purple if they show conserved differences in one or more of the 12 α-Pcdhs (see **D**).

D. α-Pcdh sequence logos of EC1:EC4 interface residues observed in the α-Pcdh structures, for each of the 12 mouse α-Pcdh isoforms, generated from sequence alignments of isoform orthologs from up to 29 species. Conserved isoform-specific residues are underlined. Secondary structure is indicated, and colors of residue numbers at top and bottom correspond to part (C).

E. Close up of the EC1:EC4 interface in the $\beta6_{EC1–4}$ (top panel, EC1 yellow, EC4 grey) and $\beta8_{EC1–4}$ (bottom panel, EC1 pink, EC4 grey) structures.

F. Surface representation of β8 EC1 and EC4 interfacial regions from an opened out dimer, highlighting the putative specificity-determining residues. Interfacial residues are colored orange if they are conserved amongst all β-Pcdhs and shades of purple if they

differ in one or more of the 22 mouse β-Pcdhs. Residues in EC1 and EC4 are colored matching shades of purple to show their predominant interaction.

G. β-Pcdh sequence logos of interface residues observed in the β-Pcdh structures are shown for a subset of the mouse β-Pcdh isoforms, for which at least 29 orthologs could be identified. The β6 and β8 interface residues are shown above the logos. Secondary structure is indicated, and colors of residue numbers at top and bottom correspond to part (F).

Figure 3. Diversity in the EC2:EC3 and EC3:EC3 interfaces of α-Pcdhs

A. The $\alpha7_{EC1–5}$ dimer structure, with the EC2:EC3 interface (red box) and the EC3:EC3 interface (dashed red circle) highlighted.

B. Close-up view of the EC2:EC3 interface in $\alpha4_{EC1–4}$ (top panel, EC2 cyan, and EC3 grey) and $\alpha7_{EC1-5}$ structures (bottom panel, EC2 in green, EC3 in grey). Side chains are shown and labeled for all residues where the side chain contributes to the dimer interface. Bound calcium ions are shown as green spheres.

C. Close-up views of the EC3:EC3 interface in the α-Pcdh dimer structures. Residue 298 from the EC3 FG-loop makes a symmetrical contact with itself in both dimers: shown in the left hand panel for α4 (cyan and grey protomers) and the right hand panel for α7 (green and grey protomers).

D. Surface representation of α7 EC2 and EC3 interfacial regions from an opened out dimer, highlighting the putative specificity-determining residues. Interfacial residues are colored orange if they are conserved amongst all α-Pcdh isoforms and shades of purple if

they differ in one or more of the 12 α-Pcdhs. Matching shades of purple denote EC2 and EC3 interacting residues.

E. α-Pcdh sequence logos of EC2 and EC3 interface residues observed in the α-Pcdh structures for each of the 12 mouse α-Pcdh isoforms, generated from sequence alignments of isoform orthologs from up to 29 species. Secondary structure is indicated, and colors of residue numbers at top and bottom correspond to part (D).

Figure 4. Diversity in the EC2:EC3 and center of symmetry interfaces of β-Pcdhs

A. The $\beta6_{EC1–4}$ dimer structure, with the EC2:EC3 interface (red box) and the center of symmetry interfaces (dashed red circle) highlighted.

B. Close-up view of the EC2:EC3 interface in $\beta6_{EC1–4}$ (top panel, EC2 yellow, and EC3 grey) and $\beta8_{EC1–4}$ structures (bottom panel, EC2 in pink, EC3 in grey). Side chains are shown and labeled for all residues where the side chain contributes to the dimer interface. Bound calcium ions are shown as green spheres.

C. Close-up views of the center of symmetry interfacial regions in the β6 (left hand panel, protomers yellow and grey) and β8 (left hand panel, protomers pink and grey) dimer structures.

D. Surface representation of β8 EC2 and EC3 interfacial regions from an opened out dimer, highlighting the putative specificity-determining residues. Interfacial residues are colored orange if they are conserved amongst all β-Pcdh isoforms and shades of purple if they differ in one or more of the 22 β-Pcdhs.

E. β-Pcdh sequence logos of key EC2 and EC3 interface residues observed in the β-Pcdh structures are shown for a subset of the mouse β-Pcdh isoforms, for which at least 29

orthologs could be identified. The β6 and β8 interface residues are shown above the logos. Secondary structure is indicated, and colors of residue numbers at top and bottom correspond to part (D).

Figure 5. Mutants between trans-interacting residues generate new α-Pcdh specificities.

A. Schematic representation of the strategy used to assess the contribution of isoform-specific conserved interface residues. (i) Wild type and mutant Pcdhs where interacting interfacial residues (shown in the diagram in EC2 and EC3) have been mutated. (ii) and (iii) Possible experimental outcomes for cell aggregation assays. When a mutant isoform prefers self-recognition to recognition of its parent wild-type isoform, separate aggregates form (panel ii). When a mutant isoform demonstrates either a similar or stronger preference to interact with the wild-type isoform than to itself, a mixed aggregate forms (panel iii).

B–E. Assessment of α7 mutants in cell aggregation assays. α7 residues were mutated to the corresponding residues in other α isoforms. Mutated-in residues are shown in red in the sequence alignments, and are displayed in red in opened-out surface representations of the α7 recognition dimer. Interfacial regions are shaded in dark gray (top and middle rows, B–D panels i and ii, and E panel i). The results of aggregation assays are shown in the bottom row. Modeling results for the two possible homophilic (wild-type:wild-type and mutant:mutant) and one possible heterophilic complexes (wild-type:mutant) are shown (B–D panel iii, and E panel ii). Models marked by green circles show favorable

interactions, and red circles indicate unfavorable interactions. The inclusion of an 'E' in the circle indicates that electrostatic interactions play a primary role.

Figure 6. Mutants between trans-interacting residues generate new β-Pcdh specificities.

A-E. Assessment of β6 mutants in cell aggregation assays. β6 residues were mutated to the corresponding residues in other β isoforms. Mutated-in residues are shown in red in the sequence alignments, and are displayed in red in opened-out surface representations of the β6 recognition dimer. Interfacial regions are shaded in dark gray (top and middle rows, A–E panel i). The results of aggregation assays are shown in the bottom row. Modeling results for the three possible complexes are shown (A–E panel ii). Models marked by green circles show favorable interactions, and red circles indicate unfavorable interactions. The inclusion of an 'E' in the circle indicates that electrostatic interactions play a primary role, whereas 'V' indicates van der Waals interactions predominate.