Clustered protocadherins (Pcdhs) mediate numerous neural patterning functions, including neuronal self-recognition and non-self-discrimination to direct self-avoidance among vertebrate neurons. Individual neurons stochastically express a subset of Pcdh isoforms, which assemble to form a stochastic repertoire of cis-dimers. We describe the structure of a Pcdh by87 cis-homodimer, which includes the membrane-proximal extracellular cadherin domains ECS and EC6. The structure is asymmetric with one molecule contributing to the interface surface from both ECS and EC6, and the other only from EC6. Structural and sequence analyses suggest that all Pcdh isoforms will dimerize through this interface. Site-directed mutants at this interface interfere with both Pcdh cis-dimerization and cell surface transport. The structure explains the known restrictions of cis-interactions of some Pcdh isoforms, including α-Pcdhs, which cannot form homodimers. The asymmetry of the interface approximately doubles the size of the recognition repertoire, and restrictions on cis-interactions among Pcdh isoforms define the limits of the Pcdh recognition unit repertoire.

clustered protocadherin | crystal structure | protein–protein interaction | neuronal self-avoidance | self-recognition

Clustered protocadherin (Pcdh) proteins are cadherin superfamily members whose genes are organized in three tandemly arranged clusters (Pedha, Pedhf, and Pedhy). Pcdhs have been implicated in many aspects of neuronal development, including olfactory, serotonergic, retinogeniculate, and cortico-cortical wiring (1–6), neuronal survival (7–11), dendrite arborization (12–15), neuronal self-avoidance (16–18), and neuronal tiling (19).

In both vertebrates and invertebrates, neuronal self-avoidance relies on the generation of unique individual cell surface identities through the stochastic expression of diverse repertoires of cell surface protein isoforms (20, 21). In Drosophila and other arthropod invertebrates, individual-neuron identities are provided by the expression of single-cell-specific subsets of Dscam protein isoforms generated by stochastic alternative splicing (22–26). By contrast, in vertebrates, the clustered Pcdhs provide analogous neuronal cell surface diversity, with single-cell-specific Pcdh isoform expression generated by stochastic promoter choice (7, 27–30).

The repertoire of Pcdh isoforms expressed in individual mammalian neurons is determined by the unique organization of the three tandem gene clusters, Pedha, Pedhf, and Pedhy (31). Each of these clusters contains multiple alternative “variable” exons (14 α, 22 β, and 22 γ in the mouse genome), which encode the full Pcdh ectodomain regions, including six extracellular cadherin (EC) domains, a single transmembrane region, and a short cytoplasmic extension. The last two variable exons in the Pedha gene cluster and the last three variable exons of the Pedhy gene cluster diverge in sequence from other Pcdh “alleles” and are referred to as “C-type” Pcdhs (αC1, αC2, γC3, γC4, and γC5) (31, 32). These C-type Pcdhs are expressed “deterministically” rather than stochastically like the non–C-type alternate α/β/γ-Pcdhs (17, 19, 28–30). Based on sequence differences, Pedhy genes have been further divided into two subfamilies—PedhyA and PedhyB (31).

Significance

Pcdhs are cell surface homophilic recognition proteins expressed stochastically to assign individual identities to each neuron. These individual identities ensure repulsion between neurites from the same cell and ensure that neurites from different cells do not repel. However, it is difficult to understand how only ~60 Pcdh isoforms can provide sufficient diversity for the billions of neurons in vertebrate nervous systems. Here, we report the crystal structure of a Pcdh cis-dimer through which individual Pcdh isoforms associate to form diverse bivalent recognition units. The structure reveals asymmetry in the cis-dimer interaction and suggests restrictions on isoform combinations compatible with cis-dimerization. These findings provide a framework to understand Pcdh cis-dimerization and the compositions of functional repertoires of Pcdh recognition units.


Reviewers: E.Y.J., University of Oxford; and E.O., University of Chicago. The authors declare no conflict of interest.

Published under the PNAS license.

Data deposition: The atomic coordinates and structure factors for the β87 EC3-6 crystal structure have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID code 5V5X).

K.M.G. and R.R. contributed equally to this work.

To whom correspondence may be addressed. Email: tm2472@cumc.columbia.edu, bh6@cumc.columbia.edu, or lss8@columbia.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1713449114/-/DCSupplemental.
between different isoforms, although there are clear exceptions to this rule. For example, α-Pcdhs and Pcdhγ4 do not homo-cis-
dimerize. Rather, they form heterodimers with members of other
subfamilies, and this is required for them to reach the cell surface
(34, 38).

Here, we present the crystal structure of a cis-dimeric membrane-
proximal Pcdhγ7EC3-6 fragment. The cis-dimer is asymmetric, with
one molecule contributing interface surface from both EC5 and
EC6, and the other only from EC6. Analysis of the putative in-
terfaces formed between different Pcdh isoforms offers a structural
explanation of the known cis-recognition properties of different
Pcdh isoforms. Structure-based sequence analysis and site-directed
mutagenesis shows that α-Pcdhs can only participate as the EC6-
only half of the interface, while “carrier” Pcdhs required for α-Pcdh
transport to the cell surface are compatible with either the EC6-
only or EC6–6 positions of the asymmetric dimer. The cis-dimeric
Pcdhγ7EC3-6 structure we present defines the overall architecture
of the canonical Pcdh recognition unit and suggests the range of
structurally allowed dimer combinations that define functional
Pcdh repertoires.

Results
Crystal Structure of the Asymmetric Pcdh Cis-Dimer. Biophysical and
mutagenesis studies have shown that Pcdhs cis-dimerize via their
EC6 and potentially EC5 domains (34, 36, 38). To structurally
characterize the cis-interface, we sought to crystallize Pcdh
fragments containing EC5 and EC6. We obtained diffracting crystals and subsequently solved the structure of a cis-dimeric
EC3–6 fragment from Pcdhγ7 at 3.5 Å resolution [dimerization disso-
ciation constant, 59.0 ± 3.4 μM (38)] (Fig. 1 and Fig. S1). The
structure was solved by molecular replacement using the
monomeric structure of γB2EC3-6 (38) as a search model. Data
collection and refinement statistics are given in Table S1.

As expected, the γB7EC3-6 protomer structure consists of four
seven-stranded EC domains connected by linkers containing three
bound calcium ions. The structure is decorated with O-mannoses
and N-linked glycans away from interfacial regions (Fig. L4). The
protomer structure is overall similar to the monomeric γB2EC3-6
structure [root-mean-square deviation (RMSD) of 2.5 Å over 408
Ca atoms; Table S2], and the EC5–6 regions of the γB2 and
γB7 structures are highly similar (RMSD of 1.2 Å over 213 Ca atoms).

The γB7EC3-6 protomer structure is more divergent overall
from the previously published monomeric γAEC3-6 structure
(RMSD of 4.7 Å over 407 Ca atoms; Table S2). However, this
is primarily due to differences in the EC4–EC5 bend angle: The
EC5–6 regions of γA4 and γB7 are very similar (RMSD of 0.8 Å
over 185 Ca atoms).

The γB7EC3-6 crystal structure contains four molecules in the
crystallographic asymmetric unit, consisting of two independent
copies of the EC3–6-mediated cis-dimer (Fig. S1). The two copies
in the crystal show essentially identical conformations, and the
interface overlaps with previously published mutations that in-
terfere with Pcdhs cis-dimerization (L555D, V560D, and R595D
γB7 numbering; Fig. S1) (38), providing functional evidence
supporting the biological relevance of the structure. Remarkably,
the cis-interface is asymmetric, with one side mediated by EC5 and
EC6 from one protomer, and the other side by EC6 alone (Fig. 1).

The same molecular surface of EC6 is utilized on both sides of the
interaction, with the EC6-only protomer contributing a few EC6
residues that are not interfacial in the EC5–6 protomer. The in-
terface comprises EC5:EC6 interactions and EC6:EC6 interactions
and buries a total of 1,644 Å2 surface area over both protomers.

The small EC5:EC6 intermolecular interface (690 Å2 buried
surface area) is formed by residues from the EC5 A, F, and G
strands and the EF loop interacting with residues from the EC6
BC and DE loops (Fig. 1C and Fig. S1). The interface consists of
both hydrophobic and hydrophilic residues centered around
W571 from the EC6 BC loop, which is conserved in all Pcdh
isoforms (Fig. S2). Above, EC6 BC loop residues G567 and
H568 make close contacts with EC5 residues R517, L519, and
L426. Below, the aliphatic regions of arginine residues R590
(EC6) and R494 (EC5) interact and the R590 side chain also
potentially forms two hydrogen bonds with the main-chain car-
bonyls of EC5 EF loop residues H490 and E491 (Fig. S1). The
larger EC6:EC6 intermolecular interface comprises residues from
the A, A′, B, and B′ strands and AB and DE loops of the
EC6-only protomer, and residues from the A, B, and B′ strands
and BC and DE loops of the EC5–6 protomer (Fig. 1C and Fig.
S1). The EC6:EC6 interactions are primarily hydrophobic, with
the center of the interface consisting of two clusters of hydro-
phobic interactions: the first among EC6-only A- and B-strand
residues L531, Y532, and V560 with EC5–6-side BC-loop resi-
dues A570 and W571, and the second involving EC6-only B-strand
residues L555 with L519 and V562 from the EC5–6 side. These
two clusters are located either side of the sole charged residue buried
in the heart of the interface, B-strand residue K558 from the EC6-only side, which likely forms hydrogen bonds with nearby mainchain carbonyls. The remaining charged EC6:EC6 interface residues are located around the periphery of the interface (Fig. 1C and Fig. S1).

The asymmetric nature of the interface implies that for a cis-dimer emanating vertically from the membrane, the EC5-6-side protomer will be positioned closer to the membrane than the EC6-only protomer (Fig. 1B). Notably, all Pcdhs have a linker of 23–25 aa between EC6 and the transmembrane helix, which is predicted to be unstructured, likely allowing flexibility in the arrangement of the dimer with respect to the membrane. We denote the EC6-only protomer as “L” (for long) because it extends further from the membrane plane, and we denote the EC5–6 protomer as “S” (for short). An important implication of the asymmetry is that there will be two distinct conformations of each cis-heterodimer; one where protomer “1” corresponds to L and where protomer “2” corresponds to S, and vice versa. To explore the mechanistic consequences of this finding, we first consider whether the asymmetric structure can account for the known cis-dimerization characteristics of the various Pcdh subfamilies. Furthermore, while there are conserved differences between different subfamilies, sequence conservation suggests that cis-dimer interfaces will be identical for Pcdh isoforms within the same subfamily.

α-Pcdhs and Pcdh γC4 can only participate as the EC6-6 side of the interface. In cell culture-based experiments, all alternate α-Pcdhs and the γC4-isomorph must form cis-heterodimers with other, “carrier” (β, γA, γB, αC2, γC3, or γC5), isoforms to reach the surface (34, 36, 43). These observations, along with mutations showing that mutational disruption of the cis-interface led to a loss of surface transport for a γ-Pcdh, led us to suggest that all Pcdh isoforms require cis-dimerization for cell surface delivery (38). As described below, the distinct cis-dimerization properties of different Pcdh subfamilies can be explained by analysis of the cis-dimer structure presented here.

Multiple sequence alignment of mouse α-Pcdhs revealed high conservation of EC5 and EC6 cis-interface residues among α-Pcdh isoforms (Fig. 2). Consensus between the sequence logo generated from multiple sequence alignment of mouse α-Pcdhs with one generated from a multiple sequence alignment of mouse carrier β-, γA-, and γB-Pcdhs revealed 11 conserved differences in interface residues between α and carrier Pcdhs (Fig. 2 B and C). Notably, a number of these residues exhibited significant differences in physicochemical characteristics between α and carrier isoforms. For example, there is a conserved α-specific arginine at position 560 in place of a carrier-Pcdh conserved valine residue (γB7 numbering: Fig. 2B).

We asked whether replacing interface residues from (β/γ) carrier Pcdhs with interface residues from α-Pcdhs would impact the cis-interactions of the carrier. Residue 560 is conserved as a valine in “carrier” Pcdhs but is conserved as arginine in α-Pcdhs. We replaced V560 with arginine in a representative of each of the carrier Pcdh subfamilies. This mutation resulted in impaired cell surface delivery of γB7 but did not prevent cell surface delivery of β17, γA9, or γB6 (Fig. 3 A and B). However, the V560R mutation was sufficient to prevent cell surface delivery of a γB7/EC6 heterodimer by β17, γA9, and γB6, although γB7 V560R still showed some carrier ability, indicated by the presence of small aggregates alongside single cells for the γB7ΔEC1 V560R with α4 cotransfection (Fig. 3C). Furthermore, addition of another mutation to an α-specific residue, Y532G, in combination with the V560R mutation, completely prevented cell surface delivery of β17 and γB6 (Fig. 3B). We also assessed the consequence of swapping in a number of other α-specific residues and observed that the double-mutant V560R/S595S also prevented cell surface delivery of γB6 (Fig. S4). To determine whether these cell surface delivery-inhibiting mutants interfered with cis-dimerization, we performed sedimentation equilibrium analytical ultracentrifugation (AUC) experiments in the context of the whole ectodomain, which revealed that the Pcdh γB6 EC1–C563R mutant (V560R in γB7 numbering) was a dimer in solution, rather than a tetramer (a dimer-of-dimers) like the wild-type molecule (Fig. 3D and Fig. S5), indicating that the mutation abrogated cis-dimerization. Additionally, introduction of the Y532G mutation to the cis-dimeric γB7 EC6–6 fragment also resulted in a disruption of cis-dimerization in solution (Fig. 3D and Fig. S5). Therefore, mutation of one or two α-specific cis-interface residues is sufficient to prevent carrier Pcdh self-delivery to the cell surface and delivery of coexpressed α-Pcdhs, similar to the behavior of singly expressed wild-type α-Pcdhs.

Importantly, because of the asymmetric nature of the cis-interface, the impact of an α-specific interface residue on the cis-interaction depends on whether it occupies the EC5–6 side (S side) or the
EC6-only side (L side) of the interface. For example, considering position 560 in carrier Pcdhs, from the EC6-only side, the side chain of this valine, V560, is central to the hydrophobic core of the interface, but the much larger and positively charged arginine, conserved in α-Pcdhs, is unlikely to be accommodated. In contrast, from the EC5–6 side, residue 560 is peripheral to the interface, and there is space for an arginine to be accommodated in the pocket it occupies. Furthermore, there is also an aspartic acid residue (544), potentially available to form a salt bridge with α-Pcdh that could not be delivered to the cell surface like γ-Pcdhs. Placing the equivalent (A570R) mutation into γB7 A570R mutant was still able to deliver α4 to the cell surface, suggesting that the mutation prevented its interaction with wild-type carriers (Fig. 4C and Fig. S6). We obtained similar results when these experiments were repeated using Pcdh γ6EC3-4 wild type (WT) and γB6EC3-4 with α-specific cis-interface mutation V563R (V560R in γB7 numbering), and also for Pcdh γB7EC3-4 wild type and γB7EC3-4 with α-specific cis-interface mutation V532G. *Dissociation constants for the monomer-to-dimer/dimer-to-tetramer transitions. †Data from ref. 38.

To test this hypothesis, we performed mutagenesis experiments designed to target each side of the interface separately. First, we targeted the EC6-only side by mutating leucine 555 to aspartic acid. L555 is present on both sides of the cis-interface; however, from the EC6-only side, it is buried in a hydrophobic pocket at the center of the EC6:EC6 portion of the interface (Fig. 4B), whereas from the EC5–6 side it is peripheral to the interface (Fig. S1). Introducing this mutation into the carrier Pcdh γB7 resulted in a loss of cell aggregation, implying that this mutation impaired cell surface delivery. This γB7 mutant was also unable to carry α4 to the cell surface (Fig. 4B). However, placing the equivalent mutation into α4 did not prevent α4 being carried to the cell surface by a wild-type carrier Pcdh (Fig. 4B and Fig. S6). Consistent results were obtained when these experiments were repeated using Pcdh γB6 and α7 and γB7 A570R mutant was also able to deliver α4 to the cell surface like γB7 A570R (Fig. S6). We obtained similar results when these experiments were repeated using Pcdh γ6B6 and α7 isoforms, although in this case γB6 A573R showed impaired self-delivery. However, γB6 A573R was still able to deliver α4 to the cell surface like γB7 A570R (Fig. S6). These data are consistent with α-Pcdhs forming the EC5–6 side of the interface in complexes with carrier Pcdhs.

To target the EC5–6 side of the γB7 cis-interface, we mutated A570 to arginine. A570 is only present on the EC5–6 side of the interface and is 100% buried (Fig. 4C and Fig. S1). Expression of γB7 with this mutation in K562 cells resulted in cell aggregation like the wild-type (Fig. 4C), implying that this mutation did not completely impair cell surface delivery. The γB7 A570R mutant was also still able to deliver α4 to the cell surface, suggesting that the mutation did not impair the ability of α4 to form a heterodimer with α-Pcdhs. Placing the equivalent (A570R) mutation into α4, however, resulted in a mutant α-Pcdh that could not be delivered to the cell surface, suggesting that the mutation prevented its interaction with wild-type carriers (Fig. 4C and Fig. S6). We obtained similar results when these experiments were repeated using Pcdh γB6 and α7 isoforms, although in this case γB6 A573R showed impaired self-delivery. However, γB6 A573R was still able to deliver α4 to the cell surface like γB7 A570R (Fig. S6). These data are consistent with the carrier Pcdhs adopting the EC6-only side of the interface when in complex with α-Pcdhs.

Like α-Pcdhs, the C-type Pcdh γC requires coexpression with a carrier Pcdh for cell surface delivery (34). To determine the molecular basis for this behavior, we carried out structure-guided sequence analysis, which revealed that, like α-Pcdhs, γC appears to have a nonfunctional EC6-only interface and a functional EC5–6 interface. Most critically, carrier-specific residues 532, 560, 568, and 590, which play a key role specifically in the EC5–6 side of the interface, have different physicochemical characteristics in carriers and γC (Fig. S3), suggesting, in analogy to α-Pcdhs, the reason why γC does not form homodimers.

γA-Pcdhs dimerize weakly but can form both sides of the cis-interface. γA-Pcdhs do not form measurable cis-homodimers in solution, and carrier Pcdhs, and α

Fig. S6 and α

α and

cis

† Data from ref. 38.
while β- and γB-Pcdhs do form cis-homodimers in solution (38). However, since γA-Pcdhs reach the cell surface when expressed alone, we postulate that they can cis-dimerize in the membrane environment (38). Sequence comparisons between interfacial residues from β- and γA-Pcdhs suggest that the weakened homophilic cis-binding characteristics of γA-Pcdhs, compared with β-Pcdhs, are likely the result of differences in the EC5–6 interface side. The γA EC6-only interfacial residues are highly conserved among γA-Pcdhs and, with the exception of residue 543, appear to be identical to β-Pcdhs interfacial residues (Fig. 2B).
Since residue 543 only contributes to the dimeric interface exclusively through main chain contacts, this residue difference is unlikely to be significant. By contrast, the EC5–6 interface is highly variable in the EC5 region among γA-Pcdhs in contrast to β- and γB-Pcdhs (Fig. 2C). Taken together, these residue differences suggest that γA- and β-Pcdh dimerization affinity differences arise primarily due to suboptimal γA EC5–6 interfaces. In addition, this observation suggests that γAs will show a preference for occupying the EC6-only side of the cis-dimer rather than the EC5–6 side in heterodimers with β-Pcdhs.

**Modeling the Interactions of Discrete Cis-Dimeric Recognition Units.** Pcdh interactions between cell membrane surfaces form the molecular basis for their neuronal patternning functions, including neuronal self-vs.-nonself discrimination. The results presented here, alongside previously published work (33, 34, 36, 38, 43), strongly suggest that Pcdhs appear on cell surfaces as both homophilic and heterophilic cis-dimers where each arm of the dimer interacts homophilically with a corresponding arm on an apposed cell surface. Two distinct models have been proposed for the intermembrane assembly formed by Pcdhs between membrane surfaces based upon their cis- and trans-interactions. One model invokes discrete dimer-of-dimer recognition complexes, and the second involves a zipper-like assembly of cis-dimers (36). The first model assumes that two identical cis-dimers on apposing cell surfaces will form a dimer-of-dimers (a cis/trans-tetramer) and that the repertoire of cis-dimers is diverse enough to account for neuronal “bar-coding.” This model encounters statistical difficulties in accounting for both self-recognition and nonself discrimination (34).

We can now, in addition, ask whether the model is structurally feasible. Fig. S4 shows a model of a γB7 EC1–6 cis-dimer constructed by superimposing the published crystal structure of γB7 EC1–4 (38) and the γB7 EC3–6 cis-dimer structure over the EC3–4 domains (RMSD = 1.3 Å over 182 Cα atoms). It is important to emphasize that no modeling is involved in its construction: its structure is completely defined by the two crystal structures.

Because of the cis-dimer asymmetry, there will be three different possible ways for two cis-dimers to associate in trans. Specifically, since dimer 1 and dimer 2 are each composed of L and S monomers, three nonequivalent structures can form: one with trans-interactions between the L protomer from dimer 1 and the S protomer from dimer 2 (Fig. 5 B, i); one with trans-interactions between L protomers from both dimers (Fig. 5 B, ii); and one with S/S trans-interactions (Fig. 5 B, iii). Each of these models was generated by simple structural superposition of the γB7 cis- and trans-dimer crystal structures over EC3–4 as in Fig. S4. The known trans-interacting interfacial residues are shown in blue and red. As is evident from the figure, none of the three models forms a dimer-of-dimers with all four of the arms engaged in trans-interactions (Fig. 5B). Since the EC1–6 regions of γB-Pcdhs form stable tetramers in solution (36, 38), this suggests that some conformational change occurs to allow both arms to engage in trans. Given the presence of flexible regions connecting EC domains, such a conformational change might well occur.

The molecular zipper model for Pcdh recognition assumes that Pcdhs dimers form a one-dimensional assembly where each arm of a Pcdh cis-dimer on one cell interacts in trans with two different cis-dimers on an apposed cell. We generated an atomic model of γB7 EC1–6 engaging simultaneously in cis- and trans-interactions using the cis-dimer structure reported here and the trans-dimer structure of γB7 EC1–4 we reported previously (38). This procedure produced a “zipper-like” lattice assembly consistent with our previous prediction (36). The zipper is regular if trans-dimers form between the EC6-only protomer (L) from one cis-dimer and the EC5–6 protomer from an apposed cis-dimer (S) (Fig. 5C). The model places the C termini of neighboring cis-dimers at a periodic interval of 106 Å (Fig. 5C).

---

**Fig. 5.** Combining the Pcdh cis- and trans-interactions to model Pcdh ectodomain interactions. (A) Structural superposition of the γB7EC1–4 crystal structure (PDB ID code 5SZP, blue) with the γB7EC3–6 cis-dimer structure (pink) over the overlapping EC3–EC4 domains generates a model of two Pcdh full-length ectodomains engaged in cis. (B) Structural superpositions of γB7EC1–4 fragments, γB7EC3–4 cis-dimers, and γB7EC1–4 trans-dimer crystal structures generate three different models of trans-interacting cis-dimers (i–iii). Transinterfacial residues are shown in blue and red. Distances between the trans-interacting residues 41 and 338 in the EC1:EC4 interface of the nonengaged arms are shown, highlighting the degree of conformational change that would need to occur for formation of a fully trans-engaged dimer-of-dimers. (C, i) A lattice assembly unit consisting of three EC1–6 cis-dimers (from Fig. 5) built by forming trans-interactions between the short protomer from one cis-dimer with the long protomer from the next by structural superposition to the γB7EC1–4 trans-dimer structure. (ii) Propagation of this lattice unit results in regular zipper-like structure, shown viewed from above (eye icon in i). (iii) Schematic of the zipper-like structure forming between apposed membrane surfaces.
The regular assembly shown in Fig. 5C assumes that all trans-interactions are between the L monomer on one cell surface and an S monomer on the other. However, if one of the arms of a dimer corresponds to an α-Pcdh, which appears to only form the S protomer in cis-dimer interactions, then formation of the strictly homophilic trans-interaction by an α-Pcdh could result in an S/S trans-interaction, which could disrupt the regularity of the lattice. While these results show that the cis-dimer structure is compatible with formation of the proposed Pcdh zipper, further work is needed to validate and understand this potential mode of interaction.

Discussion
Here, we present the crystal structure of a cis-dimeric Pcdh fragment from γB7, revealing an asymmetric cis-dimer interaction, which is likely common to all Pcdh isoforms. The dimer is formed between the EC5 and EC6 domains of one protomer and the EC6 domain of the other protomer. The mutagenesis experiments we report verify that this asymmetric dimer represents the EC6-dependent cis-interaction observed for Pcdh isoforms in solution, and is the dimer responsible for delivery of α-Pcdhs to the cell membrane by carrier Pcdhs. Structure-based sequence analysis of Pcdhs from different subfamilies reveals the molecular basis for differences in cis-interaction characteristics observed for different Pcdhs. This analysis shows that α-Pcdhs and Pcdhβ4 are incapable of forming one side of the asymmetric dimer interaction, explaining why these isoforms are unable to dimerize, and thus must form heterodimers with carrier Pcdhs to reach the cell surface.

Pcdh Cis-Dimerization and Cell Surface Expression. Cis-dimerization of Pcdhs has been shown to be required for their cell surface delivery in K562 and HEK293 cells (34, 36, 38, 43). The cis-dimer crystal structure presented here shows that our previously published mutations that prevent cell surface delivery map to the cis-dimer interface, strongly supporting this conclusion. Based on these data, we have used cell surface delivery as a proxy for assessing cis-dimerization. Specifically, we identified α-Pcdh-specific residues in the cis-interface which, when substituted into carrier Pcdhs, inhibited cis-dimerization in solution and prevented cell surface delivery of carrier Pcdhs from both the β- and γ-Pcdh families (Fig. 3). Moreover, consistent with the idea that cis-homomultimers are excluded from the same interface, additional mutations designed to disrupt the cis-interaction also prevented carrier cell surface delivery (Fig. 4).

There is currently limited data regarding Pcdh cell surface delivery in neurons although α- and carrier γ-Pcdhs have also been observed to colocalize and interact in hippocampal neurons and neuroblastoma cell lines (43). It remains unclear why Pcdhs require cis-association for cell surface localization, although our results do suggest that all Pcdhs that reach the cell membrane are properly folded, with functional cis-interfaces, and are in preformed recognition units.

The Cis-Dimer Repertoire. Although it has been previously suggested that Pcdh cis-dimer recognition units could form promiscuously among Pcdh isoforms, there are notable exceptions to this behavior, especially regarding the α- and γ-A-Pcdh gene clusters. The analysis reported here indicates that members of these subfamilies are expected to preferentially participate in one side of the interface. For illustrative purposes, we have used the results of the analysis above to predict the cis-dimer repertoire in a mature olfactory sensory neuron (OSN). Single OSN RNA-sequencing data shows that Pcdh alternate isoforms are stochastically expressed but, in contrast to Purkinje neurons, do not constitutively express C-type isoforms (17). Since these data are likely limited by insufficient depth of single-cell sequencing, we have used the cell expressing the most Pcdhs isoforms (nine) observed as our example. Fig. S7 shows the putative Pcdh cis-dimer repertoire of a mature OSN expressing three α-, three β-, and three γ-Pcdhs. Taking into account the asymmetry of the cis-interaction such that two nonequivalent cis-dimers are formed by interaction between two isoforms—i.e., β1(S)/β1(L) and β17(S)/β1(L)—the repertoire generated by nine isoforms would include 81 recognition units if all possible cis-dimers could form (double the number formed if asymmetry is not taken into account) (Fig. S7). However, the inability of α-Pcdhs to form one side of the interface limits the cell surface repertoire to 54, and the preference of the γA-Pcdh to form one side of the interface could further decrease the diversity to 48 units (Fig. S7). We note that different types of neurons may express distinct patterns of Pcdh repertoires. For example, serotonergic neurons express primarily the Pdhdh and γ-C-type isoforms, and not the Pcdhβ, β or γ “alternate” isoforms, and PcdhdhC2 alone is required for their normal tiling throughout the brain (19).

Relating Structural Models to Neuronal Diversity. We showed above that a dimer-of-dimers arrangement is consistent with structural evidence only if one allows for some degree of conformational change to allow both cis- and trans-interactions to form simultaneously. Moreover, such a dimer-of-dimers is consistent with previous observations of Pcdh ectodomains forming tetramers in solution (36, 38). The linear zipper arrangement, in contrast, emerges directly from superpositions of the available crystal structures, although such zippers do not appear to form in solution. It is therefore of interest to explore whether either model can account for the required level of neuronal diversity in a well-defined in vivo system for self-vs.-nonself discrimination. Here, we focus on the problem of OSNs assembling into a glomerulus in the olfactory bulb.

There are estimated to be 6.6–10 million OSNs in the mouse (44, 45), which coalesce to form ∼3,600 glomeruli, each consisting of OSNs that express the same olfactory receptor (46). Each glomerulus therefore contains an average ∼2,300 OSNs, suggesting that, in wild-type mice, ∼2,500,000 pairwise interactions of neurons can occur without repulsion (∼2,300 × 2,300 neurons divided by 2). If we require that the probability of inappropriate repulsion for any one of the 2,500,000 potential pairs of neurons converging to form a glomerulus is small, for example, 5 × 10−2 (implying a 95% success rate in forming glomeruli with no inappropriate repulsion), then the probability of a single pair of neurons repelling is 2 × 10−6 (SI Materials and Methods). We use this approximate number to evaluate the diversity generated by the discrete dimer-of-dimers (tetramer) model and the zipper model.

In the dimer-of-dimers model, diversity is a result of the dilution of matched isoform pairs on different cells through their incorporation into cis-dimers with isoforms that are mismatched (47). Since each neuron expresses multiple Pcdh isoforms stochastically selected from a pool of only 53 distinct isoforms, it is highly likely that two interacting neurons will randomly express one or more common isoforms. This immediately raises the question of what proportion of the expressed isoforms can be shared by two interacting cells before self/nonself recognition is impaired, that is, what is the “tolerance” for the presence of common isoforms in the two cells. Based on the results of cell aggregation assays with coexpressed Pcdh isoforms the tolerance for common isoforms is surprisingly high (34). Indeed, the results of Table S3, assuming that nine distinct isoforms are stochastically expressed per OSN, suggest that not even a single mismatched isoform can be tolerated without inappropriate repulsion. That is, only when all expressed isoforms are identical is a probability of less than 10−6 achieved. It is hard to imagine how the dimer-of-dimers model can achieve this level of specificity.

In the zipper model, incorporation of a Pcdh cis-dimer into the zipper-like assembly requires one isoform in the dimer to match the isoform on the exposed end of the zipper assembly (Fig. 5C, iii). If a matched isoform is not available, zipper assembly will terminate. Since cis-dimerization is promiscuous, the isoform on
In addition to neuronal self-vs.-nonself discrimination, Pedhs have a number of other functions in neuronal organization for which Pedh cis-association behavior, cis-dimer architecture, and the trans-interaction complexes formed by these cis-dimers will be relevant (48). For example, Pedh cell surface diversity may play a cell-autonomous role in the proper organization of dendrites in Purkinje neurons (15); additionally, Pedh trans-interactions formed between astrocytes and neurons are involved in dendrite complexity (14). Further studies are required to delineate the relationship between Pedh protein interactions and their in vivo neuronal functions.

Summary. Here, we have presented the crystal structure of a Pedh cis-dimer from Pedh yB7. Conservation analysis suggests that all Pedh isoforms will form cis-dimers through related interfaces. The formation of cis-dimers is critical to generating functional diversity for Pedhs, which are expressed as only 58 distinct isoforms, compared with 19,008 for Dscams, which mediate self-avoidance in arthropod invertebrates. With the asymmetry of the cis-dimer and the structural preferences described here, we expect that ~2,500 distinct Pedh cis-dimer recognition units can be formed. However, if recognition is between discrete Pedh recognition units (dimer-of-dimers model), the potential neuronal diversity is insufficient to account for the assembly of thousand of OSNs into a glomerulus. In contrast, the linear zipper/chain-termination model we described previously (36) solves this diversity problem, because even a single isoform mismatch is sufficient for nonself discrimination. The cis-dimer structure we report here is compatible with this chain termination model, although much work remains to acquire a full understanding of Pedh-mediated neuronal recognition.

Materials and Methods

Proteins were expressed in suspension-adapted HEK293 Freestyle cells (Invitrogen) in serum-free media and purified by nickel affinity and size exclusion chromatography. yB7 EC3–6 crystals were grown at 22 °C in 1:1 drops of 5.3 mg/mL protein with 20% ethylene glycol, 10% PEG8000, 10% Morphex amino acids additive (Molecular Dimensions), and 0.1 M Morphex Buffer System 2 (Hepes/Mops buffer; Molecular Dimensions), pH 7.5. Diffraction data were collected at 100K on beamline 24-ID-C at APS. The yB7 EC3–6 structure was solved by molecular replacement using an ensemble of all three chains from the yB2 EC3–6 structure (PDB ID code 5SZR) as the search model. Single and cotransfection K562 cell aggregation assays were performed as previously described (34).

ACKNOWLEDGMENTS. We thank George Mountoufaris for insightful discussions. We thank Surajit Banerjee for help with synchrotron data collection at the Advanced Photon Source (APS) NE-CAT 24-ID-C beamlines, supported by National Institutes of Health (NIH) Grant P41GM103403. This work was supported by National Science Foundation Grant MCB-1412472 (to B.H.) and a joint NIH grant (R01GM107571 and R01MH114817) (to T.M. and L.S.). The computing in this project was supported by two NIH instrumentation grants (S10OD012351 and S10OD021764) received by the Department of Systems Biology at Columbia University.