How clustered protocadherin binding specificity is tuned for neuronal self/non-self-recognition

Kerry M. Goodman[1,6], Phinikoula S. Katsamba[1,6], Rotem Rubinstein[5]
Göran Ahlsén[1], Fabiana Bahna[1], Seetha Mannepalli[1], Hanbin Dan[4],
Rosemary Sampogna[4], Lawrence Shapiro[1,2*], and Barry Honig[1,2,3,4*]

[1]Zuckerman Mind, Brain and Behavior Institute, Columbia University, New York, NY 10027, USA.
[2]Department of Biochemistry and Molecular Biophysics,
[3]Department of Systems Biology,
[4]Department of Medicine, Division of Nephrology, Columbia University, New York, NY 10032, USA.
[5]Department of Biochemistry, Tel Aviv University, Israel
[6]Equal contribution
[*]Correspondence: lss8@columbia.edu (L.S.); bh6@cumc.columbia.edu (B.H.)

Abstract (152 words)

The stochastic expression of fewer than 60 clustered protocadherin (cPcdh) isoforms provides diverse identities to individual vertebrate neurons and a molecular basis for self/non-self-discrimination. cPcdhs form chains mediated by alternating *cis* and *trans* interactions between apposed membranes, which has been suggested to signal self-recognition. Such a mechanism requires that cPcdh *cis* dimers form promiscuously to generate diverse recognition units, and that *trans* interactions have precise specificity so that isoform mismatches terminate chain growth. However, the extent to which cPcdh interactions fulfill these requirements has not been definitively demonstrated. Here we report biophysical experiments showing that cPcdh *cis* interactions are promiscuous, but with preferences favoring formation of heterologous *cis* dimers. *Trans*-homophilic interactions are remarkably precise, with no evidence for heterophilic interactions between different isoforms. A new C-type cPcdh crystal structure and mutagenesis data help to explain these observations. Overall, the interaction characteristics we report for cPcdhs help explain their function in neuronal self/non-self-discrimination.

Introduction

Vertebrate clustered protocadherins (cPcdhs) are arrayed in three tandem gene clusters (*Pcdhα*, *Pcdhβ*, and *Pcdhγ*), each of which encodes a number of cadherin-like proteins (Wu and Maniatis, 1999; Wu et al., 2001). cPcdhs play roles in many facets of neural development (Peek et al., 2017), including circuit development, most notably neurite self-avoidance in vertebrates (Kostadinov and Sanes, 2015; Lefebvre et al., 2012; Mountoufaris et al., 2017), and tiling (Chen et al., 2017). In self-avoidance, neurites from the same neuron (sister neurites) actively avoid one another, whereas neurons from different neurons can freely interact. Tiling is similar to self-avoidance, but in tiling all neurons acquire the same identity, so that there is uniform repulsion among self- and non-self neurites (Chen et al., 2017). Self-avoidance among sister neurites leads to the characteristic arbor structures of dendritic trees, and prevents the formation of self-synapses (Kostadinov and Sanes, 2015; Lefebvre et al., 2012).

The molecular mechanisms through which neurons discriminate self from non-self differ between vertebrate and most invertebrate animals. For arthropod invertebrates such as *Drosophila melanogaster*, self-avoidance is mediated by immunoglobulin superfamily Dscam1 cell surface proteins. The stochastic alternative splicing of *Dscam1* pre-mRNAs can, in principle, generate 19,008 distinct extracellular isoforms; the vast majority of which, based on ELISA based binding assay, mediate homophilic recognition (Miura et al., 2013; Schmucker et al., 2000; Wojtowicz et al., 2004; Wojtowicz et al., 2007). Each neuron expresses a repertoire estimated at 10–50 isoforms and the large number of Dscam1 isoforms ensures a low probability that any two contacting neurons will have an identical or even a similar isoform repertoire thus minimizing the chance of inappropriate avoidance between non-self neurons (Hattori et al., 2009).

In vertebrate nervous systems cPcdh isoform expression is controlled by the unique organization of three tandem gene clusters, *Pcdhα*, *Pcdhβ*, and *Pcdhγ* (Wu and Maniatis, 1999). Each cluster contains multiple 'alternate' exons (14 α, 22 β, and 22 γ in mouse) whose expression choices vary stochastically between cells (Canzio and Maniatis, 2019). Each gene encodes full cPcdh ectodomain regions with six extracellular cadherin (EC) domains, a single transmembrane region, and a short cytoplasmic extension (Figure 1A). The last two variable exons in the *Pcdhα* gene

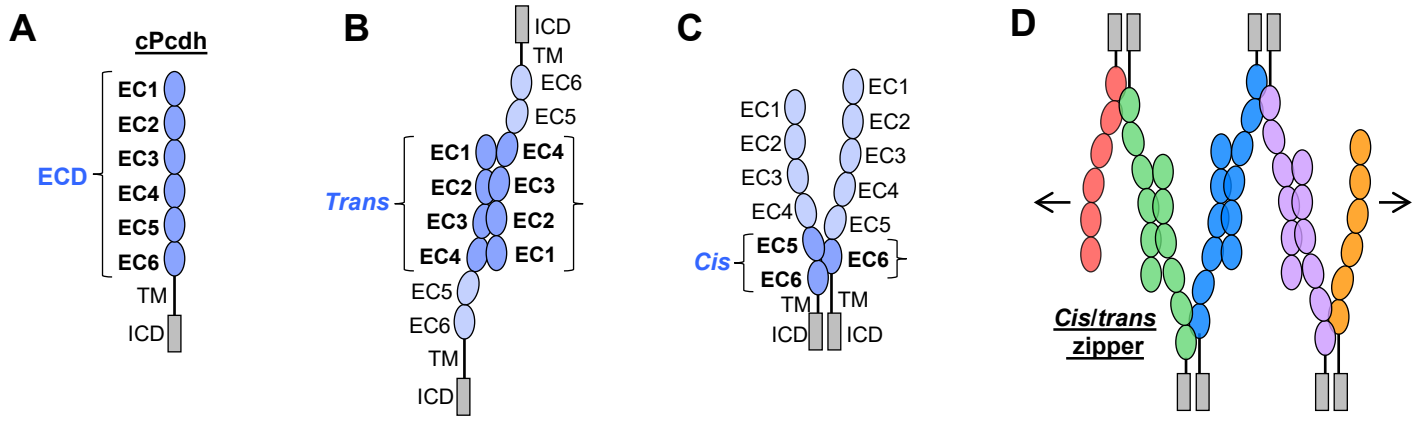


Figure 1: cPcdh domain organization and extracellular interactions
(A) Schematic depicting the domain organization of cPcdhs. EC, extracellular cadherin domain; TM, transmembrane domain; ECD, ectodomain; ICD, intracellular domain.
(B) Schematic of two cPcdhs interacting via the EC1–4 *trans* interface.
(C) Schematic of two cPcdhs interacting via the EC5–6/EC6 *cis* interface.
(D) Schematic depiction of the *cis*/*trans* cPcdh zipper comprising multiple cPcdh isoforms (various colors) engaged in homophilic *trans* interactions and promiscuous *cis* interactions as required for the proposed "isoform-mismatch chain-termination model" of cPcdh-mediated neuronal self-avoidance.

cluster and the last three variable exons of the *Pcdhγ* gene cluster diverge in sequence from other cPcdh 'alternate' isoforms and are referred to as 'C-type' cPcdhs (αC1, αC2, γC3, γC4, and γC5) (Wu and Maniatis, 1999; Wu et al., 2001). Sequence differences further subdivide *Pcdhγ* genes into two subfamilies—*PcdhγA* and *PcdhγB* (Wu and Maniatis, 1999). The *Pcdhα* and *Pcdhγ* gene clusters also contain three 'constant' exons that encode cluster-specific intracellular domains. In contrast to Dscam1, cPcdh isoform expression is controlled by alternate promoter choice although promoters controlling the expression of the C-type cPcdhs were found to be constitutively active in all Purkinje neurons (Esumi et al., 2005; Kaneko et al., 2006). While the cPcdh and Dscam1 systems bear striking similarities, the relatively small number of cPcdh isoforms – fewer than 60 – has presented a significant challenge to generation of sufficient diversity to provide vertebrate neurons with functionally unique identities.

Solution biophysics and functional mutagenesis studies, have shown that cPcdhs interact in *trans* through antiparallel interactions between their EC1–EC4 regions (Rubinstein et al. 2015), and crystal structures of alternate cPcdh *trans*-homodimers have revealed interfaces involving EC1:EC4 and EC2:EC3 (Figure 1B) (Goodman et al., 2016a; Goodman et al., 2016b; Nicoludis et al., 2016; Rubinstein et al., 2015; Thu et al., 2014). cPcdhs also form *cis* dimers through their membrane-proximal EC5–EC6 regions, and are presented on cell surfaces as *cis* dimers (Rubinstein et al., 2015; Thu et al., 2014). Crystal structures of *cis*-interacting protocadherin ectodomains (Brasch et al., 2019; Goodman et al., 2017) have revealed an asymmetrical interaction mode, where one molecule interacts through elements of EC5 and EC6, and the other interacts exclusively through EC6 (Figure 1C). To date, all structural studies of cPcdh interactions have been carried out on alternate protocadherins. Here we extend our molecular understanding of cPcdhs to C-type isoforms as well, with the goal of understanding the evolutionary design of the entire family.

In order to explain how about 60 cPcdh isoforms can provide a comparable or even greater level of neuronal diversity as 19,000 Dscam isoforms, Rubinstein et al. (2015) proposed that cPcdhs interacting located on apposed membrane surfaces would form an extended zipper-like lattice linear through alternating *cis* and *trans* interactions (Figure 1D). In self-interactions – between two membranes with identical cPcdh repertoires – these chains would grow to form large

structures, limited mainly by the number of molecules (Brasch et al., 2019; Rubinstein et al., 2015). However, in non-self-interactions – between two membranes with differing cPcdh repertoires – such large linear assemblies would not form since even a single mismatch between expressed isoforms would terminate chain assembly (Brasch et al., 2019; Rubinstein et al., 2017; Rubinstein et al., 2015). This "isoform-mismatch chain-termination model" for the "barcoding" of vertebrate neurons envisions the assembly of long cPcdh chains between sites of neurite-neurite contact to represent the signature of "self", and to be translated by downstream signaling that leads to self-avoidance (Fan et al., 2018). X-ray crystallographic studies and cryo-electron tomography (cryo-ET) studies of the full-length cPcdh ectodomains bound between the surfaces of adherent liposomes revealed the existence of linear zippers thus providing strong evidence supporting the validity of the model (Brasch et al., 2019). However, crucial questions remain unanswered. Here, a number of them are addressed.

1) For the proposed mechanism to successfully explain neuronal barcoding, *cis* interactions must be promiscuous to generate diverse repertoires of *cis*-dimeric biantennary 'interaction units', while *trans* interactions must be highly specific so that mismatched isoforms do not inappropriately enable growth of the chain through heterophilic interactions. While cell aggregation assays have suggested *trans* homophilic specificity, these assays only reflect a *competition* between different cell populations and thus don't inform as to the strength of heterophilic interactions. Moreover, the results of cell aggregation assays depend critically on the *relative* strengths of homophilic and heterophilic interactions and thus do not inform as to actual binding affinities (Honig and Shapiro, 2020). It is thus necessary to establish the extent to which heterophilic *trans* interactions are truly disallowed.

2) The assumption that *cis* interactions are promiscuous is based in large part on the fact that α-cPcdhs and γC4 cannot reach the cell surface without binding to another "carrier" isoform (Bonn et al., 2007; Goodman et al., 2016b; Murata et al., 2004; Schreiner and Weiner, 2010; Thu et al., 2014). As is the case for *trans* interactions, the strength of *cis* interactions has only been probed quantitatively in a small number of cases so that the term "promiscuous" is qualitative at best. In fact, as compared to γB and β cPcdh isoforms, most γA-Pcdhs do not form measurable *cis* homodimers in solution or, in some cases they can form weak *cis* homodimers with a $K_D > 100\mu M$,

as in the case of γA3$_{3–6}$. Nevertheless, all γA-Pcdhs are still able to reach the cell surface when expressed alone. This observation can be understood if the *cis* dimerization affinity of γA-Pcdhs is large enough to enable them to dimerize in the 2D membrane environment (Goodman et al., 2016b; Wu et al., 2013). Nevertheless, their weak dimerization affinities suggest, more generally, that cPcdhs may exhibit a range of *cis* dimerization affinities. We establish below that a wide range of affinities does in fact exist and, strikingly, most homophilic *cis* interactions are weaker than their heterophilic counterparts. We consider the functional implications of this novel observation in the discussion.

3) Structures have not yet been determined for complete C-type cPcdh ectodomains. Yet these isoforms play unique functional roles, some of which have no apparent connection to isoform diversity. For example, a single C-type isoform is sufficient for tiling which can be simply understood in terms of the formation of zippers containing identical homodimers so that all interacting neurons will avoid one another (Chen et al., 2017). Moreover, a recent paper by Garrett and coworkers discovered that neuronal survival and postnatal viability is controlled solely by γC4 suggesting a function that is unique to this isoform (although it presumably requires β and/or other γ carriers to reach the cell surface) (Garrett et al., 2019). Below we report extensive biophysical interaction studies of C-type isoform ectodomains and report the first crystal structure of a *trans* dimer formed by γC4. Our findings reveal that the specialized functions of C-type cPcdhs probably do not involve unique structural or biophysical properties of their ectodomains.

Overall, in accordance with the requirements of the isoform-mismatch chain-termination model, we find that *trans*-homophilic interactions are remarkably precise, with no evidence for heterophilic interactions between different cPcdh isoforms. In contrast cPcdh *cis* interactions are largely promiscuous but with relatively weak intra-subfamily and, especially, homophilic interactions. Possible implications of this somewhat surprising finding are considered in the discussion. Our study reveals how the extraordinary demands posed by the need to assign each neuron with a unique identity are met by an unprecedented level of protein-protein interaction specificity.

Results

cPcdh trans *interactions are strictly homophilic*

We generated biotinylated ectodomain fragments including the *trans*-interacting EC1–4 regions (Nicoludis et al., 2015; Rubinstein et al., 2015) of six representative α, β, γA, and γB cPcdh isoforms – α7, β6, β8, γA8, γA9 and γB2 – which include the most closely related isoforms by sequence identity from the β and γA subfamilies (β6/8 and γA8/9) (Rubinstein et al., 2015). These molecules were coupled over independent Neutravidin-immobilized flow cells and *trans*-interacting ectodomain fragments of multiple members of each cPcdh subfamily, including the C-types (α4, α7, α12, β6, β8, γA4, γA8, γA9, γB2, γB4, γB5, αC2, γC3, γC4, and γC5), were then flowed over the six cPcdh surfaces to assess their binding. The SPR binding profiles reveal strictly homophilic binding (Figure 2A). Remarkably, no heterophilic binding was observed for any of the analytes over any of the six surfaces (Figure 2A). Even β6/8 and γA8/9 that have 92% and 82% sequence identities respectively in their *trans*-binding EC1–4 regions exhibit no heterophilic binding. Mutations designed to disrupt α7, β6, and γA8 *trans* interaction inhibited homophilic binding, demonstrating that the observed binding occurs via the *trans* interface (Figure 2—figure supplement 1A) (Goodman et al., 2016a; Goodman et al., 2016b; Rubinstein et al., 2015). This behavior is unlike that of other adhesion receptor families where, whether they display homophilic or heterophilic preferences, the signal is never as binary as the one shown in Figure 2 (Honig and Shapiro, 2020). We estimate that, even for these pairs, the lower limit for K_D would be ~200 μM.

Much of the original evidence as to homophilic specificity was based on cell aggregation assays (Schreiner and Weiner, 2010; Thu et al., 2014) and it is of interest to compare the results obtained from these assays to those obtained from SPR. We do this in the context of examining the heterophilic binding specificity between $\beta6_{1–4}$ and $\beta8_{1–4}$ *trans* fragment that share 92% sequence identity and differ at only five residues (Figure 2—figure supplement 2A), within their respective binding interfaces. Each of these residues was mutated individually and in combination. Figure 2—figure supplement 2B and C display SPR profiles and cell aggregation images, respectively, for wild type β6 and β8 and for the various mutations. We first note that changing all five residues in β6 to those of β8 generates a mutant protein with essentially wild type β8 properties; it binds strongly to β8 but not to β6 as seen in SPR and also forms mixed aggregates with β8 but not β6.

Figure 2

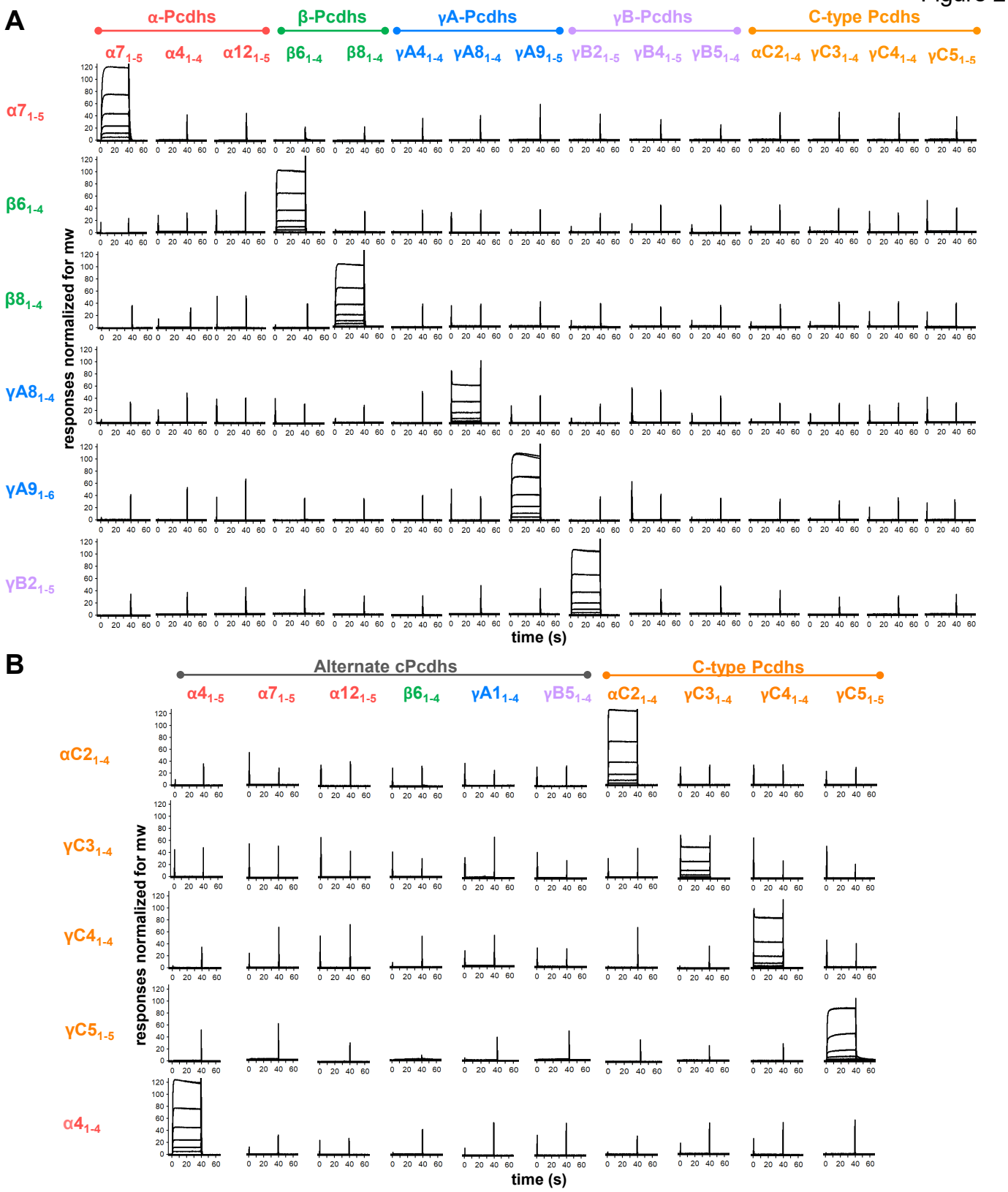


Figure 2: cPcdhs show strict homophilic specificity in their *trans* interactions

(A) SPR binding profiles of cPcdh *trans* fragment analytes from all cPcdh subfamilies (denoted in the top row) flowed over six surfaces coated with alternate cPcdh *trans* fragments (rows). Responses over all surfaces are drawn on the same scale and normalized for molecular weight.

(B) SPR binding profiles of cPcdh *trans* fragment analytes from all cPcdh subfamilies (shown in columns) flowed over individual surfaces coated with C-type and α4 cPcdh *trans* fragments (rows). Responses over all surfaces are drawn on the same scale and normalized for molecular weight.

In contrast, most of the single residue mutants retain β6-like properties in both assays whereas double and triple mutants exhibit intermediate behavior between β6 and β8. These results demonstrate that despite the 92% sequence identity between β6 and β8, their highly specific homophilic properties can be attributed to five interfacial residues. Moreover, the cell aggregation assays are consistent with the heterophilic binding traces measured by SPR; cells expressing mutants that generate strong SPR signals with either wild type β6 or β8 also form mixed aggregates with cells expressing the same wild type protein.

Of note, *trans*-interacting fragments of all four C-type cPcdhs tested showed no binding over the alternate cPcdh SPR surfaces (Figure 2B). To test whether C-type cPcdhs also show strict homophilic specificity with respect to each other we coupled biotinylated *trans*-interacting fragments of αC2, γC3, γC4, and γC5 to SPR chips and passed the same four fragments alongside alternate cPcdh *trans* fragments over these four surfaces. Only homophilic binding was observed, with each of the four C-type fragments binding to its cognate partner and no other isoform (Figure 2B). Disrupting the γC5 *trans* interaction with the S116R mutation (Rubinstein et al., 2015), inhibited binding to the γC5 surface, demonstrating that the observed binding occurs via the *trans* interface (Figure 2—figure supplement 1B).

In contrast to the other C-type isoforms, αC1 does not mediate cell-cell interactions in cell aggregation assays even when co-expressed with cPcdhs that facilitate cell-surface delivery of γC4 (Thu et al., 2014). Although we have been able to produce an αC1 EC1–4 fragment the recombinant molecule forms disulfide-linked multimers which are likely non-native, precluding confident examination of αC1's potential *trans* interactions. Notably, the sequence of mouse αC1 reveals the EC3:EC4 linker does not contain the full complement of calcium-coordinating residues, which may impact the structure and binding properties of this protein (Thu et al., 2014).

Since all the cPcdh *trans* fragment molecules used in these SPR experiments homodimerize our SPR data cannot be used to determine accurate binding affinities (Rich and Myszka, 2007). We therefore used AUC to measure the *trans*-homodimer K_Ds (Figure 2—source data 1) revealing a

wide range of binding affinities from 2.9 μM ($\alpha7_{1–5}$) to >500 μM ($\gamma C4_{1–4}$). Strict homophilic specificity is therefore not contingent on the strength of the homophilic interaction.

Crystal structure of C-type cPcdh γC4 reveals EC1–4-mediated head-to-tail* trans *dimer interaction

The biophysical properties of C-type cPcdhs pose a number of interesting questions: Despite their more divergent sequences compared with alternate cPcdhs, AUC data has confirmed that C-type cPcdhs αC2, γC3, and γC5 form *trans*-dimers using their EC1–4 domains (Goodman et al., 2016b; Rubinstein et al., 2015). However, $\gamma C4_{1–4}$ behaved as a very weak dimer in AUC (K_D > 500 μM; Figure 2—source data 1), nevertheless full-length γC4 can mediate cell aggregation when delivered to the cell surface (Thu et al., 2014). In addition, C-type isoforms have unique expression profile and function compared to alternate cPcdhs (Canzio and Maniatis, 2019; Mountoufaris et al., 2016). However, there are no published crystal structures of C-type cPcdh *trans* dimers. We therefore sought to crystallize a C-type cPcdh engaged in a *trans* interaction and obtained two distinct crystal forms of γC4 EC1–4, one at 2.4 Å resolution and the other with anisotropic diffraction at 4.6/3.9/3.5 Å resolution (Figure 3A, Figure 3—figure supplement 1A,B, Figure 3—source data 1). Both crystal structures revealed an EC1–4-mediated head-to-tail *trans* dimer, however the 2.4 Å structure had an apparently partially disrupted EC2:EC3 interface, perhaps due to the pH of the crystallization condition (pH 6) impacting the three histidines present in the EC2:EC3 interaction (Figure 3—figure supplement 1B). The 4.6/3.9/3.5 Å crystal structure appears to have a fully intact *trans* interface with a total buried surface area of 3800 $Å^2$ (Figure 3A, Figure 3—figure supplement 1B).

The intact γC4 dimer is similar overall to those of the alternate α, β, γA, and γB cPcdhs (root mean square deviation over aligned Cαs (RMSD) 3.3–4.5 Å; Figure 3—source data 2,3,4). However, by RMSD the intact γC4 dimer is even more similar to the non-clustered δ2-Pcdh *trans* dimers (RMSD 2.4–3.4 Å; Figure 3—source data 2,3,4) (Cooper et al., 2016; Harrison et al., 2020), although it is comparatively divergent from the δ1 ncPcdh1 *trans* dimer (Modak and Sotomayor, 2019). This is not particularly surprising since by sequence identity γC4 is as similar over its EC1–4 domains to the δ2 non-clustered Pcdhs as it is to the alternate and other C-type cPcdhs (~40–

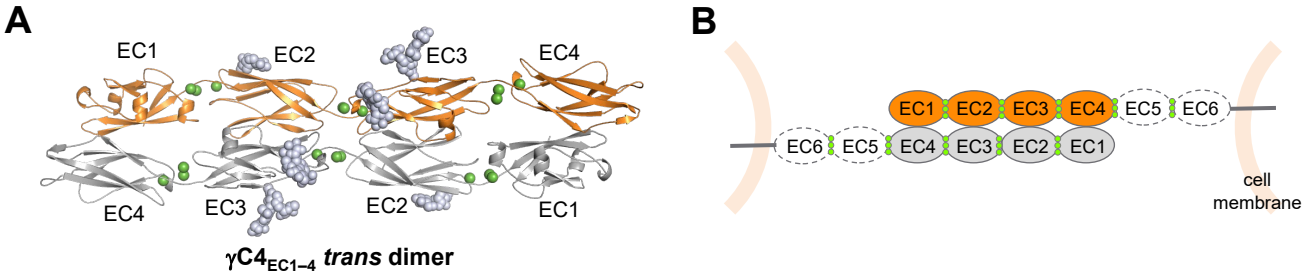


Figure 3: C-type cPcdhs adopt an EC1–4-mediated head-to-tail *trans* dimer like alternate cPcdhs
(A) Crystal structure of $\gamma C4_{EC1–4}$ reveals an antiparallel EC1–4-mediated *trans* dimer. Structure is shown in ribbon view with one protomer colored orange and the other grey. Bound calcium ions are shown as green spheres and glycans are shown as pale blue spheres.
(B) Schematic of a γC4 *trans* dimer mediating cell:cell interaction.

50% identity), with the exception of γC5 with whom it shares 54% sequence identity over EC1–4 (Figure 3—source data 5).

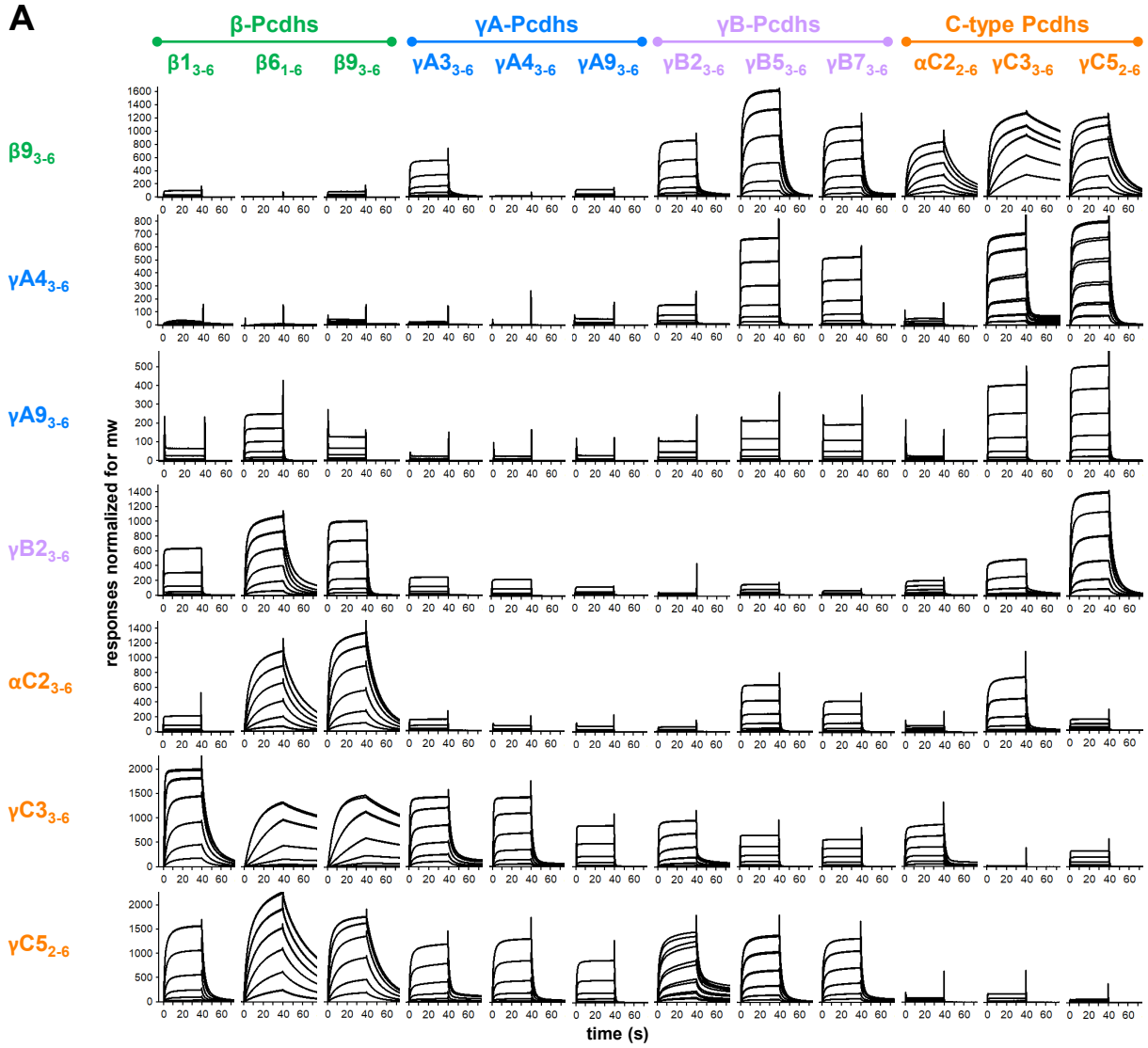
Despite the γC4 *trans* dimer sharing structural similarity and the interface having similar buried surface area as alternate cPcdhs and δ2 ncPcdhs, its binding affinity is very weak: cPcdh γB2 *trans* dimer affinity is 21.8 μM and ncPcdh19's is 0.48 μM (Harrison et al., 2020), while γC4's is >500 μM. This may be due to the presence of two buried charges in the interface, E78 and D290 (Figure 3—figure supplement 2).

Clustered protocadherin* cis *interactions are promiscuous with a range of interaction strengths

To systematically investigate cPcdh *cis* interactions, we coupled *cis*-interacting fragments of β9, γA4, γA9, γB2, αC2, γC3, and γC5 to SPR chip surfaces. *Cis*-interacting fragments of three members from each of the β, γA, and γB subfamilies (β1, β6, β9, γA3, γA4, γA9, γB2, γB5, γB7) alongside αC2, γC3, and γC5 fragments were flowed over the seven surfaces to detect their heterophilic binding (Figure 4A). α-Pcdhs, αC1, and γC4 were not included in this study since EC6-containing fragments of these molecules cannot be expressed, although an α7 EC1–5/ γC3 EC6 chimera was included among the analytes to assess the role of α7 EC5 (Figure 4—figure supplement 1C). Each of the analytes was also analyzed by AUC to determine their homophilic *cis*-interaction behavior (Figure 4—source data 1, Goodman 2016b): Four analytes, $\beta1_{3–6}$, $\gamma A4_{3–6}$, $\gamma A9_{3–6}$, and $\gamma C3_{3–6}$, are monomeric in solution as measured by AUC, therefore their SPR binding profiles could be analyzed to determine their heterophilic binding affinities (Figure 4B, Figure 4—figure supplement 1A,B). For the remaining analytes, due to the added complexity of their homophilic *cis* interactions in solution (Figure 4—source data 1) competing with their binding to the immobilized molecules, the SPR responses could not be analyzed to determine accurate K_Ds (Rich and Myszka, 2007).

The data clearly demonstrate a wide range of *cis* dimerization affinities with strong heterophilic binding signals (500–2000 RU), with much weaker homophilic binding responses typically between 100–140 RU. The strongest heterophilic *cis* interactions are in the sub-micromolar range; for example, γC3/β9 *cis*-dimerize with a K_D of 0.22 μM, while $\beta9_{3–6}$, $\alpha C2_{2–6}$ and $\gamma C5_{2–6}$ which

Figure 4



B

	K_D for heterophilic *cis* interactions (μM)			
***Cis* dimer**	**β1 3-6**	**γA4 3-6**	**γA9 3-6**	**γC3 3-6**
β9 3-6	ND	ND	ND	0.2201(2)
γA4 3-6	>50		>50	2.73(1)
γA9 3-6	>50	>50		9.60(3)
γB2 3-6	25.69(8)	44.0(6)	>50	20.00(7)
αC2 3-6	>50	>50	>50	11.51(1)
γC3 3-6	1.116(4)	3.47(2)	14.5(2)	
γC5 2-6	6.91(3)	7.74(4)	18.2(2)	>50

Figure 4: cPcdh *cis* interactions are promiscuous with a preference for interfamily heterodimers
(A) SPR binding profiles of cPcdh *cis* fragment analytes from all cPcdh subfamilies except alphas (shown in columns) flowed over individual surfaces coated with cPcdh *cis* fragments. Binding profiles for each surface are individually scaled and responses are normalized for molecular weight.
(**B**) Table of dissociation constants calculated from the SPR data for the four monomeric analytes. The number in brackets represents the error of the fit based on analysis of duplicate responses.

homodimerize with AUC-determined K_Ds of 9–35 μM. In addition to uniformly weak homophilic interactions, within-subfamily *cis* interactions were consistently among the weakest observed although a number of inter-subfamily interactions were also relatively weak (Figure 4A). For example, for the β9 surface comparatively weak binding was observed for all tested β and γA isoforms except γA3, with the monomeric β1, γA4 and γA9 producing low responses that could not be fit to a binding isotherm to calculate accurate K_Ds (Figure 4B, Figure 4—figure supplement 1B). In contrast, robust binding to the β9 surface was observed for all γB and C-type isoforms. These data are consistent with the binding responses when β9 was used as an analyte over the other six surfaces, with weak to no binding observed over the γA4 and γA9 surfaces and robust responses over the γB2, αC2, γC3, and γC5 surfaces (Figure 4A). The γA4 and γA9 surfaces showed a similar pattern of binding behaviors, with weak to no binding observed for the γA and αC2 analytes, and robust binding for the γC-cPcdhs with K_Ds for $\gamma C3_{3-6}$ of 2.73 and 9.60 μM respectively over each surface (Figure 4, Figure 4—figure supplement 1B).

Overall, these SPR data show that cPcdh *cis* binding is generally promiscuous, with measurable *cis* interactions observed for 86% of pairs tested (using a 40 RU threshold). However, the wide range of binding responses and homo- and hetero-dimeric K_Ds that span 0.2201 μM to no measurable interaction in solution suggests certain *cis* dimers will form preferentially to others. For the heterophilic binding pairs for which K_Ds could be determined (Figure 4B, Figure 4—figure supplement 1, Figure 4—figure supplement 2), the alternate cPcdhs in particular form markedly stronger *cis* heterodimers with members of different subfamilies, particularly γC3 and/or γC5, compared to their homodimeric and within-subfamily interactions. γC3 also formed stronger heterodimers with αC2 than with itself or γC5. Of note, αC2 and γC5 both form strong *cis* homodimers with K_Ds of 8.9 μM and 18.4 μM respectively as determined from AUC experiments (Figure 4—source data 1), a magnitude similar to many of their heterodimeric interactions of 11.5 μM and 6.9–18.2 μM respectively (Figure 4B).

In the next section we rationalize *cis* binding preferences in terms of the structural properties of *cis* dimers.

The asymmetric cis dimer interface and cis binding specificity

The crystal structure of the γB7 *cis* dimer revealed an asymmetric interaction, with the dimer formed by one protomer engaging using surface of both EC5 and EC6 and one protomer engaging using only EC6 (Goodman et al., 2017) with regions of EC6 overlapping in both EC5–6 and the EC6-only interfaces for all cPcdh subfamilies (Thu et al., 2014; Goodman et al., 2017). The asymmetric nature of the *cis* interaction implies that for each dimer interaction there are two possible arrangements: one with protomer "1" forming the EC5–6 side and protomer "2" forming the EC6-only side and the second where protomer "1" forms the EC6-only side and "2" the EC5–6 side. These two configurations are distinct with different residue:residue interactions. α-cPcdhs, which can only form the EC5–6 side of the *cis* dimer, require co-expression with a "carrier" cPcdh from another cPcdh subfamily, which can form the EC6-only side of the *cis* dimer, for robust delivery to the cell surface (Thu et al., 2014, Goodman et al., 2017). Although α-cPcdhs and γC4, which also requires a carrier for delivery to the cell surface, are likely to be extreme cases, sequence analysis alongside the low homodimerization ability of many cPcdh isoforms suggests many cPcdhs will more readily form one side of the *cis* interface than the other (Goodman et al., 2017).

We previously suggested that γA-cPcdhs will prefer to form the EC6-only side of the interface since they have a poorly conserved EC5 interface and do not form strong homodimers in solution (Figure 4—source data 1) (Goodman et al., 2017). The C-type cPcdh γC3 also does not form *cis* homodimers in solution. However, as shown in Figure 4, γA-cPcdhs form strong heterodimers with γC3 with dissociation constants in the low-micromolar range (Figure 4B and Figure 4—figure supplement 1B). Structure-guided sequence analysis for the γA4/γC3 dimer in both EC6-only and EC5–6 possible orientations, using the available crystal structures of the γB7 *cis* dimer and monomeric γA4 EC3–6 (Figure 5A and Figure 5—figure supplement 1), suggests that γC3 prefers to form the EC5–6 side: γC3 has a number of residue differences in interface residues that are conserved among β, γA and γB cPcdhs (V/L555, R/K558, W/V562, and S/R595) that seem likely to disfavor the EC6-only side of the interface and favor the EC5–6 side (Figure 5—figure supplement 1B,C). Two of these residues, V555 and S595, result in a potential loss of EC6-only interface buried surface area and are shared with α-cPcdhs, which cannot occupy the EC6-only

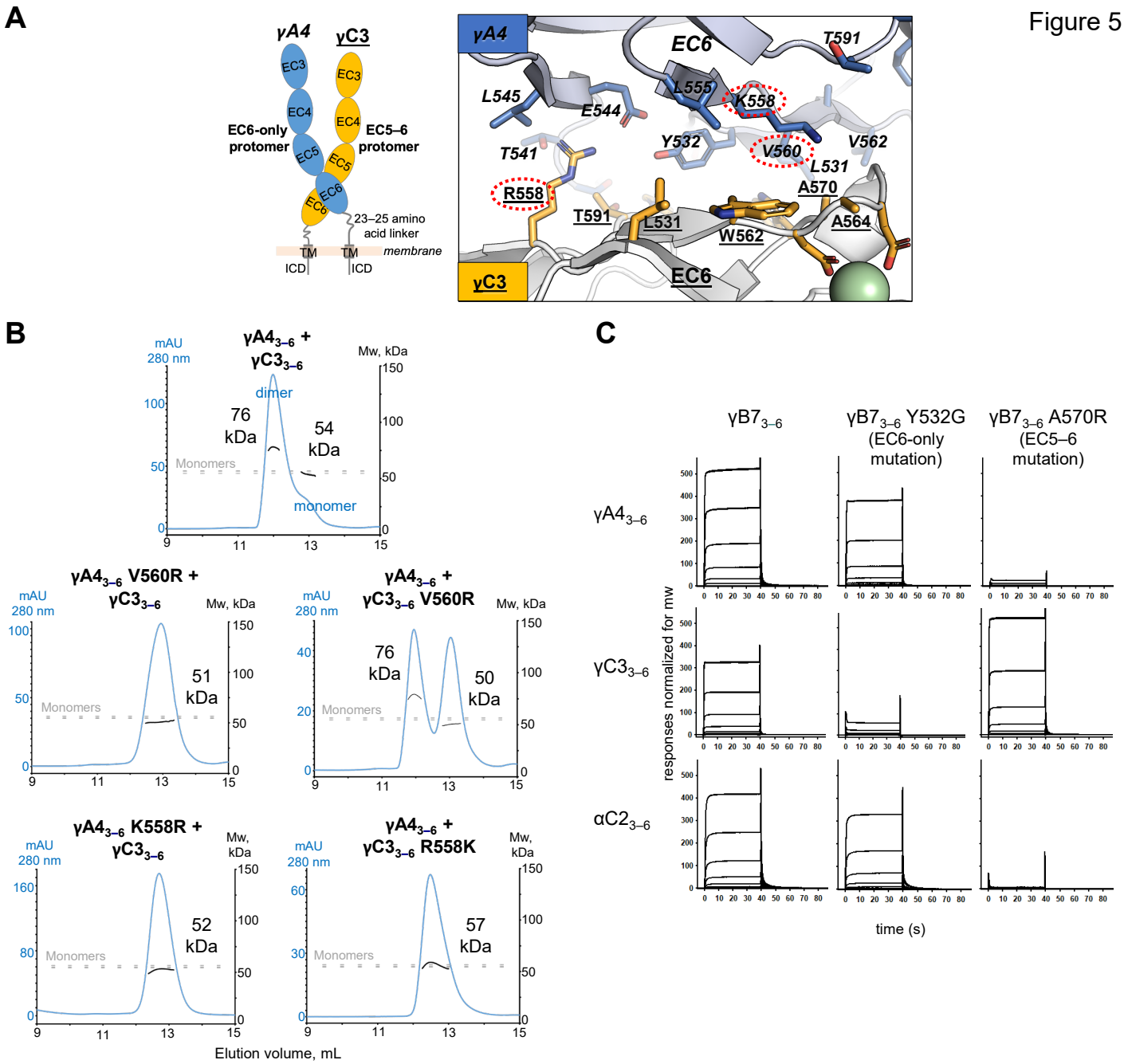


Figure 5: γA4 preferentially forms the EC6-only side and γC3 the EC5–6 side in *cis* dimers

(**A**) Structural model of γA4/γC3 *cis* dimer based on γB7$_{EC3–6}$ *cis* dimer and γA4$_{EC3–6}$ crystal structures (PDBs: 5V5X and 5SZQ). γA4 is shown adopting the EC6-only side (blue protomer) and γC3 is shown adopting the EC5–6 side (yellow protomer). Left, schematic of the γA4/γC3 EC3–6 *cis* dimer. Right, close-up view of the EC6:EC6 interface from the modeled *cis* dimer showing interfacial residue side chains. Bound calcium ions are shown as green spheres. Residues which were mutated in the panel B are circled in red. γB7 crystal structure numbering is used for both γA4 and γC3 residues. See methods for γA4 and γC3 alignment.

(**B**) *Top*, SEC-MALS data for an equimolar mixture of wild-type γA4$_{EC3–6}$ and γC3$_{EC3–6}$ showing dimer formation. Plot shows size exclusion absorbance at 280 nm trace (left axis), molecular weight of the eluant peaks (right axis), and the monomer molecular weights of γA4$_{EC3–6}$ and γC3$_{EC3–6}$ measured by mass spectrometry – 54.5 kDa and 56.5 kDa respectively – as dashed grey lines. Average molecular weight of the molecules in the dimer and monomer eluant peaks are labeled. *Middle*, SEC-MALS data for V560R mutants, which target the EC6-only side of the interface. *Bottom*, SEC-MALS data for residue 558 mutants. The γC3-like K558R mutation in γA4 inhibits heterodimer formation with wild-type γC3. Similarly, the γA4-like R558K in γC3 inhibits dimerization with wild-type γA4.

(C) SPR binding profiles for γB7$_{EC3–6}$ wild type and *cis* interface mutants flowed over three individual wild-type *cis* fragment surfaces. The two mutations specifically target one side of the *cis* interface.

position (Goodman et al., 2017). Structural analysis further suggests that γC3-specific residue R558 would not be well accommodated from the EC6-only side, potentially causing van der Waals clashes (Figure 5—figure supplement 1C). By contrast, from the EC5–6 side R558 is positioned to form an additional salt bridge with γA4 residue E544 and a hydrogen bond with Y532, promoting dimer formation (Figure 5A; Figure 5—figure supplement 1B). γA4 residue E544 is positioned to form this salt bridge due to the EC6 A/A' loop region adopting a different arrangement in the γA4 crystal structure to that observed for γB2 and γB7 in their respective crystal structures (Goodman et al., 2016c; Goodman et al., 2017).

Based on our analysis, we generated mutants of both γA4 and γC3 targeting the EC6-only side of the interface and used size exclusion-coupled multi-angle light scattering (SEC-MALS) to assess their preferred orientation on γA4/γC3 heterodimerization. In SEC-MALS wild type γA4 EC3–6 and γC3 EC3–6 behave as monomers when run alone, and form a dimer when mixed in equimolar amounts (Figure 5B; Figure 5—figure supplement 2A). The V560R mutation (γB7 numbering, see methods for sequence alignment) is based on EC6-only impaired α-cPcdhs, and has been previously shown to block γB6's homophilic *cis* interaction in solution (Goodman et al., 2017). γA4 V560R did not dimerize with wild-type γC3 whereas γC3 V560R could still dimerize with wild type γA4 (Figure 5B). Therefore impairing γA4s EC6-only interface blocks γA4/γC3 dimer formation while impairing γC3s EC6-only interface does not (although the dimerization appears to be weaker compared to the wild type γA4/γC3 *cis* interacting pairs). We also generated a γC3-like mutant of γA4, K558R, which also targets the EC6-only interface. Like γA4 V560R, γA4 K558R also did not dimerize with wild type γC3 in MALS and when replicated in SPR experiments (Figure 5B, Figure 5—figure supplement 2B). The reverse mutation in γC3, R558K, inhibited dimerization with wild type γA4 (Figure 5B). Therefore, like the α-specific R560 residue, γC3-specific R558 has distinct effects on dimerization when in γA4 or γC3, inhibiting heterodimerization when mutated into γA4 but promoting heterodimerization in γC3. Together these data suggest that the γA4/γC3 dimer has a preferred orientation, with γA4 predominantly occupying the EC6-only position and γC3 the EC5–6 side. Our data also account for the fact that

neither isoform homodimerizes in solution since the EC5–6 side would be impaired in the γA4 homodimer while the EC6 side would be impaired in the γC3 homodimer.

Next, we sought to test whether γA4 and γC3 preferentially adopt these specific positions in *cis* interactions with a γB isoform. To accomplish this we generated mutants of γB7 individually targeting the EC6-only interaction surface, γB7 Y532G, and the EC5–6 side, γB7 A570R, respectively (Goodman et al., 2017) (Figure 4—source data 1). In SPR, γB7 Y532G had only a small impact on γA4 binding, while γB7 A570R abolished γA4 binding (Figure 5C). In contrast, γB7 Y532G prevented γC3 binding while γB7 A570R showed robust γC3 binding (Figure 5C). These results suggest that γA4/γB7 and γC3/γB7 *cis* heterodimers also have preferred orientations with γA4 and γC3 maintaining their preferences for the EC6-only and EC5–6 positions respectively. Additionally, SPR data for the γB7 mutants over the αC2 surface suggests αC2 preferentially occupies the EC6-only side in αC2/γB7 dimers (Figure 5C). This is notable since αC2 forms robust *cis*-homodimers and therefore, like γB7, can presumably readily occupy both positions in its homophilic interactions, implying that the αC2/γB7 orientation preference could be specific to the particular heterodimer pairing.

Discussion

Trans specificity - The results of this study add to our current understanding of cPcdhs in a number of ways. First, they reveal a remarkable level of specificity in *trans* homophilic interactions in the sense that in no case was a heterophilic interaction detected in our SPR measurements. Prior data has clearly indicated that cPcdhs exhibit a preference for homophilic *trans* interactions but the extent of this specificity was not established in quantitative terms but were, rather, based on cell aggregation experiments. The SPR experiments with cPcdhs reported here show no evidence of cross-interaction between non-identical cPcdh isoforms. This level of specificity is unusual for cell-cell recognition proteins, as significant intra-family interactions are evident in most other families examined to date including type I cadherins (Katsamba et al., 2009; Vendome et al., 2014), type II cadherins (Brasch et al., 2018), DIPs and Dprs (Cosmanescu et al., 2018), sidekicks (Goodman et al., 2016c), and nectins (Harrison et al., 2012). Even the non-clustered delta

protocadherins, which are preferentially homophilic and utilize an antiparallel EC1–4 interface like the cPcdhs, show heterophilic intra-family *trans* interactions, though they show no cross-reactivity with cPcdhs (Harrison et al., 2020).

High fidelity homophilic interaction is a strict requirement of the chain termination model for the barcoding of vertebrate neurons and has been accomplished through the exploitation of a multidomain interface of almost 4000 $Å^2$ (Nicoludis et al., 2019) that enables the positioning of enough "negative constraints" (Sergeeva et al., 2020) to preclude the dimerization of about 1600 heterophilic pairs of 58 mouse cPcdh isoforms (Rubinstein et al., 2017). Dscams accomplish the same task for thousands of isoforms but this is accomplished by exploiting the combinatorics made possible by a three-domain interface where each domain interacts largely independently with an identical domain on its interacting partner (see discussion in (Zipursky and Grueber, 2013)). Although it is likely that Dscams dimerize with a comparable level of homophilic specificity to that of cPcdhs, the evidence is based on a semi-quantitative ELISA-type assay of recombinant multimerized isoforms (Wojtowicz et al., 2007) and analytical ultracentrifugation experiments on a few select isoforms (Wu et al., 2012).

Cis interactions – Despite early evidence that *cis* interactions are promiscuous, the data reported here indicate that this generalization needs to be significantly refined. It has already been established that α cPcdhs don't form intra-subtype *cis* interactions and can only reach the cell surface when mediated by heterophilic *cis* interactions with members of other subtype families. The data presented in Figure 4 indicate that this is an extreme example of quite general behavior: intra-subtype *cis* interactions are invariably weaker than inter-subtype interactions. We have attributed the fact that monomeric cPcdhs can reach the cell surface on their own despite not exhibiting measurable *cis* interactions to their presence on the restricted 2D surface of membranes whereas biophysical experiments are carried out in a 3D solution environment. There may of course be other, still undetermined, factors involved.

The *cis* binding preferences indicated by our data can be largely understood in terms of the asymmetric interface discussed above. Specifically, different isoforms preferentially form one side of the *cis* dimer: for example, the EC6-only side for Pcdh-γA4 and the EC5–6 side for Pcdh-γC3.

Homodimerization requires participation of single isoform on both sides of an interface posing challenges in the optimization of binding affinities since, in some cases, the same residue must participate in different intermolecular interactions. Given significant sequence conservation in all members of an alternate cPcdh subfamily (Figure 4—figure supplement 3) even intra-subfamily heterophilic interactions are more difficult to optimize relative to inter-subfamily heterodimerization where there are no constraints on the two interacting surfaces. Additionally, the robust cell surface delivery of many cPcdhs in cells expressing only a single isoform also suggests that all carrier isoforms – β-, γA-, and γB-cPcdhs, plus C-types αC2, γC3, and γC5 – can fill both the EC6 and EC5–6 roles, as *cis*-dimer formation is thought to be required for cell surface export (Goodman et al., 2017; Goodman et al., 2016b; Thu et al., 2014).

Functional implications of cPcdh structure – The functional role of precise *trans* homophilic specificity in ensuring high fidelity discrimination between neuron self and non-self has been discussed previously and summarized above. It is an essential feature of the chain termination model. The role of promiscuous *cis* interactions can also be understood in terms of this model in that *cis* promiscuity enables the formation of a large and diverse set of *cis* dimers that can only form long molecular zippers when all isoforms are matched. However, the results of this study indicate reveal strong preferences for inter-subgroup heterophilic interactions whose biological rationale is by no means clear. cPcdhs from the three subfamilies have been shown to act cooperatively in certain neuronal contexts although whether this relates to their *cis* interactions is unknown (Hasegawa et al., 2016; Ing-Esteves et al., 2018).

One possible advantage of weak homophilic *cis* interactions would be to ensure that once reaching the cell surface a diverse set of *cis* dimers forms. This explanation implicitly assumes that most isoforms (except for α-Pcdhs and γC4) reach the surface as homodimers that must then quickly dissociate and form more stable heterodimers. Another possible explanation posits that zippers consisting of homodimers are easier to form in a kinetic sense than heterodimeric zippers, and that this would reduce the diversity required in the chain termination model since, in this model, it is essential that all isoforms be in incorporated into a growing zipper. The formation of long homophilic zippers might lead to a repulsive phenotype even when mismatches are present.

However, these explanations would not easily account for interfamily heterophilic preferences. Our results for C-type isoforms suggest that other factors may be involved. C-type cPcdhs have different functions than alternate cPcdhs and these are reflected in different expression patterns. For example, αC2 can be alone responsible for tiling (Chen et al., 2017) and γC4 has a unique and crucial role in neuronal survival (Garrett et al., 2019). One possibility is that C-types form different within family extracellular interactions than alternate cPcdhs but our structural and biophysical data indicate that this is not the case and suggest the possibility that different intracellular interactions account for their specialized functions. The cytoplasmic domain plays an important role in the activation of Wnt, WAVE, and other signaling cascades (Chen et al., 2009; Fukuda et al., 2008; Keeler et al., 2015; Mah and Weiner, 2017; Onouchi et al., 2015; Pancho et al., 2020). In some cases, the cytoplasmic domains of a subset or even a single cPcdh isoform activates a specific signaling cascade. For example, cPcdh-γC3 is the only isoform able to interact and inhibit Axin1, a Wnt pathway activator (Mah et al., 2016). Of note, γ-cPcdh intracellular domains consists of a C-terminal constant region common to all γ isoforms and a membrane-proximal variable region consisting of ~100 residues that could account for the unique intracellular interactions and signaling of individual isoforms. Additionally it is possible that extracellular interactions to molecules from other families, such as Neuroligins, may account for some distinctions in function (Molumby et al., 2017; Steffen et al., 2021).

Another possibility is suggested by the observation that C-types are often highly expressed compared to alternate cPcdhs, for example in Purkinje cells (Esumi et al., 2005; Kaneko et al., 2006). To ensure sufficient diversity in growing zippers, it would then be important to ensure that zippers that are formed are not overly enriched in C-type isoforms as would be accomplished through preferential heterophilic interactions. This same logic would pertain to alternate cPcdhs as well in cases where one subfamily is more heavily expressed than another.

Overall, the results of this study demonstrate the remarkable tuning of the interactions among clustered protocadherin family members: homophilic *trans* interactions are remarkably specific despite the high level of sequence identity among family members while *cis* interactions, though somewhat promiscuous, also appear designed to have binding preferences of still uncertain function. These binding properties match requirements of the "isoform-mismatch chain-

termination model" for neuronal self-vs-non-self discrimination in which all expressed cPcdh isoforms assemble into intercellular zippers formed by alternating promiscuous *cis* and matched *trans* interactions with assembly size dictated by the presence or absence of mismatched isoforms. It remains to be seen whether such assemblies can be observed *in vivo* and how they control downstream signaling pathways.

Materials and Methods

Protein production

cDNAs for cPcdh ectodomain fragments, excluding the predicted signal sequences, were cloned into a pαSHP-H mammalian expression vector (a kind gift from Daniel J. Leahy, John Hopkins University) modified with the human Binding immunoglobulin protein (BiP; MKLSLVAAMLLLLSAARA) signal sequence and a C-terminal octa-histidine tag(Rubinstein et al., 2015). The signal sequences were predicted using the SignalP 4.0 server (Petersen et al., 2011). Point mutations were introduced into cDNA constructs using the KOD hot start polymerase (Novagen) following the standard Quikchange protocol (Stratagene).

Suspension-adapted HEK293 Freestyle cells (Invitrogen) in serum free media (Invitrogen) grown and maintained at 37 °C and 10% carbon dioxide were used for protein expression. The plasmid constructs were transfected into cells using polyethyleneimine (Polysciences Inc.) (Baldi et al., 2012). Media was supplemented with 10 mM calcium chloride 4 h after transfection. Conditioned media was harvested ~6 days after transfection and the secreted proteins were purified using batch nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography followed by size exclusion chromatography over Superdex 200 26/60 column (Cytiva) on an AKTA pure fast protein liquid chromatography system (Cytiva). Purified proteins were concentrated to >2 mg/ml in 10 mM Tris pH 8.0, 150 mM sodium chloride, 3 mM calcium chloride, and 100–250 mM imidazole pH 8.0 and stored at 4 °C for short-term use or flash frozen in liquid nitrogen for long-term storage at –80 °C.

Constructs encoding biotinylated cPcdh fragments for immobilization in surface plasmon resonance (SPR) experiments were prepared by insertion of an Avi-tag (GLNDIFEAQKIEWHE)-encoding sequence between the octa-histidine tag and stop codon. These were co-transfected with a plasmid encoding the biotin-Ligase BirA from E. coli (Lys2–Lys321) with a BiP signal sequence and a C-terminal endoplasmic reticulum-retention signal (DYKDEL) (Barat and Wu, 2007). Expression plasmid and BirA plasmid were mixed at a 9:1 ratio for transfection and 50 μM Biotin (Sigma) was added to the media 4 h post-transfection. Purification was carried out exactly as for

the non-biotinylated constructs and biotinylation was confirmed by western blot using NeutrAvidin-HRP (ThermoFisher).

Sedimentation equilibrium analytical ultracentrifugation (AUC)

Protein	Imidazole pH 8.0 (mM)	Spin speeds (rpm)
α4 EC1–5	100	9000, 11000, 13000, 15000
α7 EC1–5 L301R	100	9000, 11000, 13000, 15000
α12 EC1–5 *(poorly behaved)*	200	11000, 14000, 17000, 20000
γB4 EC1–5	200	11000, 14000, 17000, 20000
γB5 EC1–4-AVI	200	11000, 14000, 17000, 20000
γC4 EC1–4	200	15000, 18000, 21000, 24000
γC5 EC1–5 S116R	200	11000, 14000, 17000, 20000
β6 EC1–4	100	9000, 11000, 13000, 15000
β6 EC1–4-AVI	200	11000, 14000, 17000, 20000
β6 EC1–4 R41N	200	11000, 14000, 17000, 20000
β6 EC1–4 S117I	200	11000, 14000, 17000, 20000
β6 EC1–4 L125P	200	11000, 14000, 17000, 20000
β6 EC1–4 E369K	200	11000, 14000, 17000, 20000
β6 EC1–4 Y371F	200	11000, 14000, 17000, 20000
β6 EC1–4 R41N/S117I *(precipitates)*	200	11000, 14000, 17000, 20000
β6 EC1–4 R41N/E369K	200	11000, 14000, 17000, 20000
β6 EC1–4 S117I/L125P	200	11000, 14000, 17000, 20000
β6 EC1–4 R41N/S117I/L125P	200	11000, 14000, 17000, 20000
β6 EC1–4 R41N/S117I/E369K	200	11000, 14000, 17000, 20000
β6 EC1–4 R41N/E369K/Y371F	200	11000, 14000, 17000, 20000
β6 EC1–4 R41N/S117I/L125P/ E369K/Y371F	200	11000, 14000, 17000, 20000
β1 EC3–6	200	12000, 16000, 20000, 24000
β6 EC1–6	250	9000, 11000, 13000, 15000
β9 EC3–6	200	11000, 14000, 17000, 20000
γA3 EC3–6	200	11000, 14000, 17000, 20000
γA9 EC3–6	200	11000, 14000, 17000, 20000
γB7 EC3–6 A570R	200	13000, 17000, 21000, 25000
αC2 EC3–6-AVI	200	11000, 14000, 17000, 20000
γC5 EC2–6	250	9000, 11000, 13000, 15000

Experiments were performed in a Beckman XL-A/I analytical ultracentrifuge (Beckman-Coulter, Palo Alto CA, USA), utilizing six-cell centerpieces with straight walls, 12 mm path length and sapphire windows. Protein samples were dialyzed overnight and then diluted in 10 mM Tris pH 8.0, 150 mM NaCl, 3 mM $CaCl_2$ with varying concentration of imidazole pH 8.0, as detailed in the above table. The samples were diluted to an absorbance at 10 mm and 280 nm of 0.65, 0.43 and 0.23 in channels A, B and C, respectively. The dilution buffer was used as blank. The samples were run in duplicate at four speeds as detailed in the above table. The lowest speed was held for

20 h then four scans were conducted with 1 h interval, the subsequent three speeds were each held for 10 h followed by four scans with 1h interval. Measurements were taken at 25 °C, and detection was by UV at 280 nm. Solvent density and protein v-bar at both temperatures were determined using the program SednTerp (Alliance Protein Laboratories, Corte Cancion, Thousand Oaks, CA, USA). The molecular weight of each protomer used in AUC experiments, was determined by MALDI mass spectrometry. For calculation of dimeric K_D and apparent molecular weight, all data were used in a global fit, using the program HeteroAnalysis, (www.biotech.uconn.edu/auf). Calculation of the tetramer K_ds was done with the program Sedphat (http://www.analyticalultracentrifugation.com/sedphat/index.htm).

Surface plasmon resonance (SPR) binding experiments

SPR binding experiments were performed using a Biacore T100 biosensor equipped with a Series S CM4 sensor chip, immobilized with NeutrAvidin over all four flow cells. NeutrAvidin immobilization was performed in HBS-P buffer, pH 7.4 at 32°C, over all four surfaces using amine-coupling chemistry as described in Katsamba et al. (2009), resulting in approximately 10,000 RU of NeutrAvidin immobilized (Katsamba et al., 2009). Binding experiments were performed at 25°C in a running buffer containing 10 mM Tris-Cl pH 8.0, 150 mM NaCl, 3 mM $CaCl_2$, 20 mM imidazole, 0.25 mg/mL BSA and 0.005% (v/v) Tween-20 unless otherwise noted.

C-terminal biotinylated fragments were tethered over individual NeutrAvidin-immobilized flow cells (shown in the left column of each Figures 2, 4, 5C, Figure 2—figure supplement 1, Figure 4—figure supplement 1 and Figure 5—figure supplement 2) at 2300–3000 RU, depending on the experiment, using a flow rate of 20 µL/min. A NeutrAvidin-immobilized flow cell was used as a reference in each experiment to subtract bulk refractive index changes. The analytes tested in each experiment are listed at the top row. All analytes (with exceptions for the *cis* interacting pairs $\gamma C3_{3-6}/\beta 9_{3-6}$, in both orientations, and $\beta 6_{1-6}/\gamma C3_{3-6}$ in Figure 4A, discussed below) were tested at six concentrations ranging between 24, 8, 2.667, 0.889, 0.296, and 0.099 µM, similarly prepared using a three-fold dilution series. $\gamma C3_{3-6}$ binding over $\beta 9_{3-6}$ (Figure 4A) was tested at five concentrations from 8–0.099 µM.

For all experiments, analyte samples were injected over the captured surfaces at 50 μL/min for 40 s, followed by 180 s of dissociation phase, a running buffer wash step and a buffer injection at 100 μL/min for 60 s. Protein samples were tested in order of increasing concentration, and the entire concentration series was repeated again to confirm reproducibility. Every three binding cycles, buffer was used as an analyte instead of a protein sample to double reference the binding responses by removing systematic noise and instrument drift. The resulting binding curves were normalized for molecular weight differences according to data provided by mass spec for each molecule. The data was processed using Scrubber 2.0 (BioLogic Software). To provide an estimate of the number of possible heterophilic binding pairs, we have used a cut-off of 40RU, which is the lowest signal that can be observed for a homodimeric *cis* fragment pair, $\gamma B2_{3-6}$.

In Figure 4A, $\beta 6_{1-6}$ and $\beta 9_{3-6}$ were tested over $\gamma C3_{3-6}$ at six concentrations ranging from 900 to 3.7 nM, which is 27-fold lower than the other interactions, prepared using a three-fold dilution series in a running buffer containing increased concentrations of imidazole (100 mM) and BSA (0.5 mg/mL) to minimize nonspecific interactions. For these two interactions, although analyte samples were injected over the captured surfaces at 50 μL/min for 40s, the dissociation phase was monitored for 300s to provide additional time for complex dissociation. Nevertheless, higher analyte concentrations produced binding profiles that were not reproducible, most likely due to the fact that the bound complexes could not dissociate completely at these concentrations.

For the calculation of heterophilic K_Ds for the monomeric *cis* fragments $\beta 1_{3-6}$, $\gamma A4_{3-6}$, $\gamma A9_{3-6}$ and $\gamma C3_{3-6}$ over each of the six surfaces, except $\beta 9_{3-6}$, the duplicate binding responses were fit globally, using an 1:1 interaction model and a single K_D was calculated as the analyte concentration that would yield 0.5 R_{max} and a fitting error, indicated in brackets. K_Ds < 24 μM were calculated using an independent R_{max}. For K_Ds > 24 μM, the R_{max} was fixed to a global value determined by the R_{max} of a different cPcdh analyte tested over the same surface during the same experiment that showed binding above 50% and therefore produced a more accurate R_{max}. For K_Ds >50 μM, a lower limit is listed since at the analyte concentrations used we could not measure accurate K_Ds even when the R_{max} is fixed. The binding curves of $\gamma C3_{3-6}$ over the $\beta 9_{3-6}$ did not come to equilibrium during the time-course of the experiment so a kinetic analysis was performed to calculate a K_D (Figure 4—figure supplement 1A). Binding of $\gamma C3_{3-6}$ was tested using a

concentration range of 900–0.411 nM prepared using a three-fold dilution series in a running buffer containing increased concentrations or imidazole (100 mM) and BSA (0.5 mg/mL) to minimize any nonspecific interactions. Protein samples were injected over the captured surfaces at 50µL/min for 90s, followed by 420s of dissociation phase, a running buffer wash step and a buffer injection at 100 µL/min for 60s. Protein samples were tested in order of increasing concentration in triplicate to confirm reproducibility. Every three binding cycles, buffer was used as an analyte instead of a protein sample to double reference the binding responses by removing systematic noise and instrument drift. The binding data was analyzed using an 1:1 interaction model to calculate the kinetic parameters and the K_D.

K562 cell aggregation assays

Full-length Pcdhs β6 and β8 cDNAs were cloned into the pMax expression vectors encoding C-terminal mCherry or mVenus tagged Pcdh proteins, then transfected into K562 cells (ATCC CCL243) as previously described (Goodman et al., 2017; Thu et al., 2014). Point mutants were generated using the QuikChange method (Stratagene). In brief, K562 cells were cultured at 37°C with 5% CO_2 in DMEM with GlutaMAX (GIBCO) supplemented with 10% FBS and 1% penicillin-streptomycin for two days. Next, cells were counted, centrifuged, and resuspended at a density of ~1.5×10^4 cells/µL in SF Cell Line 4D-Nucleofector Solution SF with supplement according to manufacturer instructions (Lonza). 2 µg of each Pcdh expression construct were transfected into 20 µL of the K562 cell suspension by electroporation using an Amaxa 4D-Nucleofector (Lonza). Transfected cells were transferred to a 24-well plate in 500 µL of medium per well and incubated overnight at 37°C and 5% CO_2. Cells then were mixed, re-incubated with gentle rocking for 4 hours, then imaged with an Olympus IX73 fluorescent microscope to determine the extent of aggregation.

Size-exclusion coupled multi-angle light scattering (SEC–MALS)

SEC-MALS experiments were performed using a Superdex 200 Increase 3.2/300 size exclusion column on an AKTA FPLC system (GE healthcare) coupled to inline static light scattering (Dawn Heleos II, Wyatt Technology), differential refractive index (Optilab rEX, Wyatt Technology) and UV detection. Purified cPcdh proteins were diluted to 18 µM in running buffer (150 mM NaCl, 10

mM Tris-Cl pH 8, 3 mM $CaCl_2$, 200 mM Imidazole pH 8) and 50 or 100 µl samples were run at a flow rate of 0.5 ml/min at room temperature. Mixtures of cPcdh fragments were prepared in the same buffer at final concentrations of 18 µM for each protein and run under the same conditions. Data were analyzed using ASTRA software (Wyatt Technologies).

X-ray crystallography

Crystallization screening of $\gamma C4_{1–4}$ using the vapor diffusion method yielded two protein crystal forms: The first crystal form crystals were grown using a protein concentration of 7 mg/ml in 10% (w/v) PEG8000, 20% ethylene glycol, 10% Morpheus Amino Acids (Molecular Dimensions), and 0.1 M Morpheus Buffer System 2 (Hepes/MOPS buffer; Molecular Dimensions) pH 7.5. No additional cryoprotection was required for this crystal form. The second crystal form crystals were grown using a protein concentration of 7 mg/ml in 1 M LiCl, 0.1 M Mes pH 6.0, and 10% (w/v) PEG6000. The crystal used for data collection was cryo-protected in the crystallization condition plus 30% (w/v) glycerol. X-ray diffraction data for each crystal form were collected at 100K from single crystals at Northeastern Collaborative Access Team (NE-CAT) beamline 24ID-E at the Advanced Photon Source, Argonne National Laboratory.

$\gamma C4_{1–4}$ crystal form 1: Diffraction anisotropy and pseudosymmetry

The X-ray diffraction data for the first crystal form showed strong diffraction anisotropy, with relatively strong diffraction along c* and much weaker diffraction along a* and b* (Figure 3—figure supplement 1). These data were therefore truncated using ellipsoidal limits with using a 3.0 F/sigma cut-off along each of the three principle crystal axes as implemented in the UCLA Diffraction Anisotropy Server (Strong et al., 2006) to 4.6/3.9/3.5 Å. The completeness within the applied ellipsoidal resolution limits was 96.8% (Figure 3—source data 1).

$\gamma C4_{1–4}$ crystal form 1: Crystal structure phasing and refinement

The $\gamma C4_{1–4}$ crystal structure was solved by molecular replacement using Phaser (McCoy et al., 2007), implemented in ccp4 (Winn et al., 2011). The $\gamma C5_{EC1–3}$ crystal structure (PDB: 4ZPO) modified using a sequence alignment to γC4 with Phenix's MRage program (Liebschner et al., 2019) was used as a search model. Following an initial round of rigid body refinement in Phenix

(Liebschner et al., 2019) the EC domain 4 from the $\alpha7_{EC1–5}$ crystal structure (PDB: 5DZV) was manually placed into the electron density map, using structural alignment to the EC1–3 regions as a guide. The resulting model was subjected to a further round of rigid body refinement. At this stage there was clear difference density for the interdomain calcium ions and covalently linked glycans not present in the models. Iterative model building using Coot (Emsley et al., 2010) and maximum-likelihood refinement using Phenix (Liebschner et al., 2019) was subsequently conducted. The higher resolution (2.4 Å) crystal form 2 crystal structure (see below) was used as a reference model in later rounds of iterative model-building and refinement to guide the local geometry choices in this lower resolution structure. Final refinement statistics are given in Figure 3—source data 1.

$\gamma C4_{1–4}$ crystal form 2: data processing, phasing, and refinement

The $\gamma C4_{1–4}$ crystal form 2 dataset was indexed using XDS (Kabsch, 2010) and scaled using AIMLESS (Evans and Murshudov, 2013). The data was spherically truncated with high resolution limit of 2.4 Å. Data collection statistics are given in Figure 3—source data 1.

The $\gamma C4_{1–4}$ crystal form 2 crystal structure has two molecules in the asymmetric unit was solved by molecular replacement using Phaser (McCoy et al., 2007), implemented in Phenix (Liebschner et al., 2019), using the EC2–3 portion of the *trans*-dimer from the crystal form 1 crystal structure early in refinement as a search model. The molecular replacement solution was then subjected to an initial round of rigid body refinement using Phenix, followed by two rounds of model building in Coot (Emsley et al., 2010) and maximum likelihood refinement in Phenix. The two EC4 domains were then manually placed in the electron density and subjected to rigid body refinement. Following a further two iterative rounds of model building and refinement the two EC1 domains were manually placed. Iterative model-building and refinement continued yielding the final crystal structure whose statistics are given in Figure 3—source data 1.

Structure analysis

Buried surface areas were calculated using 'Protein interfaces, surfaces and assemblies' service (PISA) at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) (Krissinel and Henrick, 2007) and are given as the change in accessible surface area over both

protomers. Root mean square deviations over aligned Cα atoms (RMSDs) between structures were calculated using Pymol (Schrödinger, LLC). Crystal structure figures were made using Pymol (Schrödinger, LLC).

Sequence analysis

Multiple sequence alignments were generated using Clustal Omega (Sievers et al., 2011) and visualized using ESPript3.0 (Robert and Gouet, 2014). Sequence logos were generated from multiple sequence alignments using WebLogo3 (Crooks et al., 2004).

Amino acid sequence alignment of cPcdhs γB7, γA4, and γC3 EC1–6 regions

```
CLUSTAL O(1.2.4) multiple sequence alignment

γB7      -QPVRYSIPEELDRGSVVGKLAKDLGLSVLEVSARKLRVS--AEKLHFSVDSESGDLLVK        57
γA4      -EQIRYSVPEELERGSVVGNLAADLGLEPGKLAERGVRIVSRGKTQLFALNPRSGSLVTA        59
γC3      STIIHYEILEERERGFPVGNVVTDLGLDLGSLSARRLRVVSGASRRFFEVNWETGEMFVN        60
            ::*.: ** :**  **::. ****.  .:: * :*:   ..   * :: .:*.:..

γB7      DRIDREQICKGRRKCELQLEAVLENPLNIFHVVVEIEDVNDHAPQFPKDEINLEISESDS        117
γA4      GRVDREGLCDRSPKCTANLEILLEDKVRILAIEVEIIDVNDNAPSFGAQQREIKVAESEN        119
γC3      DRLDREELCGTLPSCTVTLELVVENPLELFSAEVVVQDINDNNPSFPTGEMKLEISEALA        120
         .*:*** :*    .*   ** ::*: :.::   * : *:**: *.*   : :::::*:

γB7      PGARTILESAKDLDIGMNSLSKYQLSPNDYFLLLVKDNPDGSKYPELELQKMLDREAEST        177
γA4      PGTRFPLPEAFDLDIGVNALQGYQLSSNDHFSLDVQSGPDGIKYPELVLENALDREEEAV        179
γC3      PGTRFPLESAHDPDVGSNSLQTYELSHNEYFALRVQTREDGTKYAELVLERALDWEREPS        180
         **:*  * .* * *:* *:*. *:** *::* * *:   ** ** ** *:. ** * *

γB7      HHLMLTAVDGGDPPRTGTTQLRIRVVDANDNRPVFSQDVYRVRLPEDLPPGTTVLRLKAM        237
γA4      HHLVLTAFDGGDPVRSGTATIQVTLVDTNDNAPVFTQPEYHISVKENLPVGTRLLTIKAT        239
γC3      VQLVLTALDGGTPARSATLPIRITVLDANDNAPAFNQSLYRARVREDAPPGTRVAQVLAT        240
          :*:***.*** * *:.*  ::: ::*:*** *.*.*  *:  : *: * ** :  : *

γB7      DQDEGINAEFTYSFLGV-ANK--AQFSLDPITGDIVTRQSLDFEEVEQYTIDVEAKDRGS        294
γA4      DPDEGVNGEVTYSFRNV-REKISQLFQLNSLTGDITVLGELDYEDSGFYDVDVEAHDGPG        298
γC3      DLDEGLNGEIVYSFGSHNRAGVRELFALDLVTGVLTIKGRLDFEDTKLHEIYIQAKDKGA        300
         * ***:*.*..*** .         * *: :** :.    **:*:   : : ::*:*  .

γB7      --LSSQCKVIIEVLDENDNRPEIIITSLSDQISEDSPSGTVVALFKVRDRDSGENAEVMC        352
γA4      --LRARSKVLVTVLDVNDNAPEVTVTSLTSSIQEASSPGTVIALFNVHDSDSGENGLVTC        356
γC3      NPEGAHCKVLVEVVDVNDNAPEITVTSVYSPVPEDAPLGTVIALLSVTDLDAGENGLVTC        360
             ::.**:: *:* *** **: :**: . : * :  ***:**:.* * *:***. * *

γB7      SLSGNNPFKIHSSSNNYYKLVTDSILDREQTPGYNVTITATDRGKPPLSSSTTITLNVAD        412
γA4      SIPDNLPFRLEKTYGNYHRLLIHRTLDREEVSDYNITITATDQGTPPLSTETYISLQVVD        416
γC3      EVPPGLPFSLTSSLKNYFTLKTSAALDRETMPEYNLSITARDSGIPSLSALTTVKVQVSD        420
         .:  . ** : .:  **. *     ****    **::*** * * * **: * :.::* *

γB7      VNDNAPVFQQQAYLINVAENNQPGTSITQVKAWDPDVGSNGLVSYSIIASDLEPKALSSF        472
```

```
γA4      INDNPPTFTHASYSAYIPENNPRGASILSITAQDPDSGENAQVIYSLSEDTIQGAPMSSY      476
γC3      INDNPPQSSQSSYDVYVEENNLPGVPILNLSVWDPDAPPNARLSFFLLEPGAETGLVSRY      480
         :*** *   : :*   : ***  *. * .:.. ***   *. : : :     :   :* :

γB7      VSVNQDSGVVYAQRAFDHEQIRSFQLTLQARDQGSPALSANVSMRVLVDDRNDNAPRVLY      532
γA4      VSINSNTGVLYALRSFDYEQFQDLKLLVTARDSGTPPLSSNVSLSLSVLDQNDNTPEILY      536
γC3      FTINRDNGVLTTLVPLDYEDQREFQLTAHINDGGTPVLATNISVNVFVTDRNDNAPQVLY      540
         .::* :.**: :   :*:*: :.::*    .* *:* *::*:*: : * *:***:*.:**

γB7      PTLEPDGSALFDMVPRAAEPGYLVTKVVAVDADSGHNAWLSYHVLQASDPGLFSLGLRTG      592
γA4      PTIPTDGSTGVELTPRSADPGYLVTKVVAVDKDSGQNAWLSYRLLKASEPGLFSVGLHTG      596
γC3      PR---PGQSSVEMLPRGTAAGHVVSRVVGWDADAGHNAWLSYSLLGAPNQSLFAVGLHTG      597
         *     *.: .:: **.:  *::*::**. * *:*:****** :* * : .**::**:**

γB7      EVRTARALSDKDAARQRLLVAVRDGGQPPLSATATLLLVFADSLQE     638
γA4      EVRTARALLDRDALKQSLVVTVQDHGQPPLSATVTLTIAVSDNIPD     642
γC3      QISTARPIQDTDSPRQILTVLISDSGEPLLSTTATLTVSVTEESPE     643
         :: *** : * *: :* * * : * *:* **:*.** : .::.  :
```

Structure-based sequence analysis of the γA4/γC3 interaction

Since both γA4$_{3–6}$ and γC3$_{3–6}$ are monomeric in solution but form a robust heterodimer when mixed (in SPR, AUC, and SEC-MALS) we hypothesized that these molecules might have opposing *cis* interaction side preferences. To facilitate hypothesis generation on the nature of their *cis* heterodimer interaction we modeled the two possible γA4/γC3 *cis* dimers: one with γA4 occupying the EC6-only position and γC3 the EC5–6 position; and the second with γC3 in the EC6-only position and γA4 in the EC5–6 position. To do this the monomeric γA4$_{3–6}$ crystal structure (PDB: 5SZQ) was structurally superimposed over EC6 domains with the EC6-only protomer from the γB7 EC3–6 *cis*-dimer crystal structure (PDB: 5V5X; RMSD 0.7 Å over 91 aligned Cαs) or over EC5–6 domains with the EC5–6 protomer (RMSD 1.0 Å over 194 aligned Cαs). Since γA4 and γB7 are so structurally similar in their EC5–6 regions modeling γA4's *cis* interactions in this manner as a basis for hypothesis generation seemed reasonable. The only region of significant structural deviation within the EC5–6 regions between γA4 and γB7 is in the EC6 A–A' loop region which has a peripheral role in the EC6-only protomer interface. The γB7 structure was used directly as a guide for our structure-based sequence analysis of γC3's *cis* interactions.

1

2 ***Cis interface mutants***

3 Our studies of Pcdh *cis* interactions we have found that mutagenesis of the *cis* interface commonly
4 has a deleterious impact on protein expression levels in our system (Goodman et al., 2017). We
5 assume this is because *cis* interaction is required for robust cell-surface delivery/secretion (Thu et
6 al., 2014), although this hasn't been specifically addressed in our HEK293 protein expression
7 system.

8

9 To test our structure-guided hypotheses regarding γA4 and γC3s' *cis* interactions and side
10 preferences as we tried to make a number of different *cis* interface mutants and were able to obtain
11 four different mutants (see table below). Since protein yields were generally too low for AUC and
12 SPR, MALS was used to study the impact of these mutants on γA4/γC3 *cis* dimer formation.

13

Mutant protein (γB7 numbering given in parentheses)	***Cis* interface side targeted**	**Protein expression in 25 mL test**
γC3 EC3–6 Y540G (Y532G equivalent)	EC6-only	No
γC3 EC3–6 V560D (L555D equivalent)	EC6-only	No
γC3 EC3–6 V565R (V560R equivalent)	EC6-only	Yes
γC3 EC3–6 A575R (A570R equivalent)	EC5–6	No
γC3 EC3–6 R563K (K558R equivalent)	Both	Yes
γA4 EC3–6 Y536G (Y532G equivalent)	EC6-only	No
γA4 EC3–6 L559D (L555D equivalent)	EC6-only	No
γA4 EC3–6 V564R (V560R equivalent)	EC6-only	Yes
γA4 EC3–6 A574R (A570R equivalent)	EC5–6	No
γA4 EC3–6 K562R (K558R equivalent)	EC6-only	Yes
β1 EC3–6 V563R (V560R equivalent)	EC6-only	No
β1 EC3–6 S573R (A570R equivalent)	EC6-only	No
β1 EC3–6 K561R (K558R equivalent)	EC5–6	No
β9 EC3–6 V563R (V560R equivalent)	EC6-only	No
β9 EC3–6 A573R (A570R equivalent)	EC6-only	No
β9 EC3–6 K561R (K558R equivalent)	EC5–6	No

14

Accession numbers

Atomic coordinates and structure factors for the γC4 EC1–4 crystal structures are deposited in the protein data bank with accession codes PDB: 7JGZ and 7RGF.

Author contributions

K.M.G., B.H., and L.S. designed experiments and analyzed data. S.M., F.B., and K.M.G. cloned, expressed, purified and crystallized the proteins. K.M.G. determined the crystal structures and conducted the sequence and structural analysis. S.M. and F.B. performed the site-directed mutagenesis. P.S.K. performed and analyzed the surface plasmon resonance experiments. G.A. performed and analyzed the analytical ultracentrifugation and multi-angle light scattering experiments. H.D. and R.S. performed and analyzed the cell aggregation experiments. K.M.G. and P.S.K. produced the figures. K.M.G. and L.S. drafted the manuscript. K.M.G., P.S.K., R.R., B.H., and L.S. edited the manuscript.

Acknowledgements

We thank Surajit Banerjee for help with synchrotron data collection at the APS NE-CAT 24-ID-C/E beamlines, supported by NIH P41GM103403. This work was supported by a joint NIH grant to Tom Maniatis, Columbia University and L.S. (R01MH114817) and a National Science Foundation grant to B.H. (MCB-1412472) .

Competing Interests

The authors declare no competing interests.

Supplementary Files

Figure 2—figure supplements 1–2 and source data 1
Figure 3—figure supplements 1–2 and source data 1–5
Figure 4—figure supplements 1–3 and source data 1
Figure 5—figure supplements 1–2

References

Baldi, L., Hacker, D.L., Meerschman, C., and Wurm, F.M. (2012). Large-Scale Transfection of Mammalian Cells. In Protein Expression in Mammalian Cells: Methods and Protocols, J.L. Hartley, ed. (Totowa, NJ: Humana Press), pp. 13-26.

Barat, B., and Wu, A.M. (2007). Metabolic biotinylation of recombinant antibody by biotin ligase retained in the endoplasmic reticulum. Biomolecular Engineering *24*, 283-291.

Bonn, S., Seeburg, P.H., and Schwarz, M.K. (2007). Combinatorial Expression of α- and γ-Protocadherins Alters Their Presenilin-Dependent Processing. Molecular and Cellular Biology *27*, 4121-4132.

Brasch, J., Goodman, K.M., Noble, A.J., Rapp, M., Mannepalli, S., Bahna, F., Dandey, V.P., Bepler, T., Berger, B., Maniatis, T.*, et al.* (2019). Visualization of clustered protocadherin neuronal self-recognition complexes. Nature *569*, 280-283.

Brasch, J., Katsamba, P.S., Harrison, O.J., Ahlsén, G., Troyanovsky, R.B., Indra, I., Kaczynska, A., Kaeser, B., Troyanovsky, S., Honig, B.*, et al.* (2018). Homophilic and Heterophilic Interactions of Type II Cadherins Identify Specificity Groups Underlying Cell-Adhesive Behavior. Cell Reports *23*, 1840-1852.

Canzio, D., and Maniatis, T. (2019). The generation of a protocadherin cell-surface recognition code for neural circuit assembly. Current Opinion in Neurobiology *59*, 213-220.

Chen, J., Lu, Y., Meng, S., Han, M.-H., Lin, C., and Wang, X. (2009). α- and γ-Protocadherins Negatively Regulate PYK2. Journal of Biological Chemistry *284*, 2880-2890.

Chen, W.V., Nwakeze, C.L., Denny, C.A., O'Keeffe, S., Rieger, M.A., Mountoufaris, G., Kirner, A., Dougherty, J.D., Hen, R., Wu, Q.*, et al.* (2017). Pcdhαc2 is required for axonal tiling and assembly of serotonergic circuitries in mice. Science *356*, 406-411.

Cooper, S.R., Jontes, J.D., and Sotomayor, M. (2016). Structural determinants of adhesion by Protocadherin-19 and implications for its role in epilepsy. eLife *5*, e18529.

Cosmanescu, F., Katsamba, P.S., Sergeeva, A.P., Ahlsen, G., Patel, S.D., Brewer, J.J., Tan, L., Xu, S., Xiao, Q., Nagarkar-Jaiswal, S.*, et al.* (2018). Neuron-Subtype-Specific Expression, Interaction Affinities, and Specificity Determinants of DIP/Dpr Cell Recognition Proteins. Neuron *100*, 1385-1400.e1386.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence Logo Generator. Genome Research *14*, 1188-1190.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallographica Section D: Biological Crystallography *66*, 486-501.

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. Nature Genetics *37*, 171-176.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallographica Section D: Biological Crystallography *69*, 1204-1214.

Fan, L., Lu, Y., Shen, X., Shao, H., Suo, L., and Wu, Q. (2018). Alpha protocadherins and Pyk2 kinase regulate cortical neuron migration and cytoskeletal dynamics via Rac1 GTPase and WAVE complex in mice. eLife *7*, e35242.

Fukuda, E., Hamada, S., Hasegawa, S., Katori, S., Sanbo, M., Miyakawa, T., Yamamoto, T., Yamamoto, H., Hirabayashi, T., and Yagi, T. (2008). Down-regulation of protocadherin-α A isoforms in mice changes contextual fear conditioning and spatial working memory. European Journal of Neuroscience *28*, 1362-1376.

Garrett, A.M., Bosch, P.J., Steffen, D.M., Fuller, L.C., Marcucci, C.G., Koch, A.A., Bais, P., Weiner, J.A., and Burgess, R.W. (2019). CRISPR/Cas9 interrogation of the mouse Pcdhg gene cluster reveals a crucial isoform-specific role for Pcdhgc4. PLOS Genetics *15*, e1008554.

Goodman, K.M., Rubinstein, R., Dan, H., Bahna, F., Mannepalli, S., Ahlsén, G., Aye Thu, C., Sampogna, R.V., Maniatis, T., Honig, B.*, et al.* (2017). Protocadherin *cis*-dimer architecture and recognition unit diversity. Proceedings of the National Academy of Sciences *114*, E9829-E9837.

Goodman, K.M., Rubinstein, R., Thu, C.A., Bahna, F., Mannepalli, S., Ahlsen, G., Rittenhouse, C., Maniatis, T., Honig, B., and Shapiro, L. (2016a). Structural Basis of Diverse Homophilic Recognition by Clustered alpha- and beta-Protocadherins. Neuron *90*, 709-723.

Goodman, K.M., Rubinstein, R., Thu, C.A., Mannepalli, S., Bahna, F., Ahlsén, G., Rittenhouse, C., Maniatis, T., Honig, B., and Shapiro, L. (2016b). γ-Protocadherin structural diversity and functional implications. eLife *5*, e20930.

Goodman, K.M., Yamagata, M., Jin, X., Mannepalli, S., Katsamba, P.S., Ahlsén, G., Sergeeva, A.P., Honig, B., Sanes, J.R., and Shapiro, L. (2016c). Molecular basis of sidekick-mediated cell-cell adhesion and specificity. eLife *5*, e19058.

Harrison, O.J., Brasch, J., Katsamba, P.S., Ahlsen, G., Noble, A.J., Dan, H., Sampogna, R.V., Potter, C.S., Carragher, B., Honig, B.*, et al.* (2020). Family-wide Structural and Biophysical Analysis of Binding Interactions among Non-clustered δ-Protocadherins. Cell Reports *30*, 2655-2671.e2657.

Harrison, O.J., Vendome, J., Brasch, J., Jin, X., Hong, S., Katsamba, P.S., Ahlsen, G., Troyanovsky, R.B., Troyanovsky, S.M., Honig, B.*, et al.* (2012). Nectin ectodomain structures reveal a canonical adhesive interface. Nature Structural & Molecular Biology *19*, 906-915.

Hasegawa, S., Kumagai, M., Hagihara, M., Nishimaru, H., Hirano, K., Kaneko, R., Okayama, A., Hirayama, T., Sanbo, M., and Hirabayashi, M. (2016). Distinct and Cooperative Functions for the

Protocadherin-α,-β and-γ Clusters in Neuronal Survival and Axon Targeting. Frontiers in Molecular Neuroscience *9*, 155.

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-U687.

Honig, B., and Shapiro, L. (2020). Adhesion Protein Structure, Molecular Affinities, and Principles of Cell-Cell Recognition. Cell *181*, 520-535.

Ing-Esteves, S., Kostadinov, D., Marocha, J., Sing, A.D., Joseph, K.S., Laboulaye, M.A., Sanes, J.R., and Lefebvre, J.L. (2018). Combinatorial Effects of Alpha- and Gamma-Protocadherins on Neuronal Survival and Dendritic Self-Avoidance. The Journal of Neuroscience *38*, 2713-2729.

Kabsch, W. (2010). XDS. Acta Crystallographica Section D: Biological Crystallography *66*, 125-132.

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. Journal of Biological Chemistry *281*, 30551-30560.

Katsamba, P., Carroll, K., Ahlsen, G., Bahna, F., Vendome, J., Posy, S., Rajebhosale, M., Price, S., Jessell, T.M., Ben-Shaul, A.*, et al.* (2009). Linking molecular affinity and cellular specificity in cadherin-mediated adhesion. Proceedings of the National Academy of Sciences *106*, 11594-11599.

Keeler, A.B., Schreiner, D., and Weiner, J.A. (2015). Protein Kinase C Phosphorylation of a γ-Protocadherin C-terminal Lipid Binding Domain Regulates Focal Adhesion Kinase Inhibition and Dendrite Arborization. Journal of Biological Chemistry *290*, 20674-20686.

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

Krissinel, E., and Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. Journal of Molecular Biology *372*, 774-797.

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.

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J.*, et al.* (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallographica Section D *75*, 861-877.

Mah, K.M., Houston, D.W., and Weiner, J.A. (2016). The gamma-Protocadherin-C3 isoform inhibits canonical Wnt signalling by binding to and stabilizing Axin1 at the membrane. Scientific Reports *6*.

Mah, K.M., and Weiner, J.A. (2017). Regulation of Wnt signaling by protocadherins. Seminars in Cell & Developmental Biology *69*, 158-171.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. Journal of Applied Crystallography *40*, 658-674.

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.

Modak, D., and Sotomayor, M. (2019). Identification of an adhesive interface for the non-clustered δ1 protocadherin-1 involved in respiratory diseases. Communications Biology *2*, 354.

Molumby, M.J., Anderson, R.M., Newbold, D.J., Koblesky, N.K., Garrett, A.M., Schreiner, D., Radley, J.J., and Weiner, J.A. (2017). γ-Protocadherins Interact with Neuroligin-1 and Negatively Regulate Dendritic Spine Morphogenesis. Cell Reports *18*, 2702-2714.

Mountoufaris, G., Canzio, D., Nwakeze, C.L., Chen, W.V., and Maniatis, T. (2016). Writing, Reading, and Translating the Clustered Protocadherin Cell Surface Recognition Code for Neural Circuit Assembly. Annual Review of Cell and Developmental Biology *34*, 471-493.

Mountoufaris, G., Chen, W.V., Hirabayashi, Y., O'Keeffe, S., Chevee, M., Nwakeze, C.L., Polleux, F., and Maniatis, T. (2017). Multicluster Pcdh diversity is required for mouse olfactory neural circuit assembly. Science *356*, 411-414.

Murata, Y., Hamada, S., Morishita, H., Mutoh, T., and Yagi, T. (2004). Interaction with Protocadherin-γ Regulates the Cell Surface Expression of Protocadherin-α. Journal of Biological Chemistry *279*, 49508-49516.

Nicoludis, J.M., Green, A.G., Walujkar, S., May, E.J., Sotomayor, M., Marks, D.S., and Gaudet, R. (2019). Interaction specificity of clustered protocadherins inferred from sequence covariation and structural analysis. Proceedings of the National Academy of Sciences *116*, 17825-17830.

Nicoludis, J.M., Lau, S.-Y., Schaerfe, C.P.I., Marks, D.S., Weihofen, W.A., and Gaudet, R. (2015). Structure and Sequence Analyses of Clustered Protocadherins Reveal Antiparallel Interactions that Mediate Homophilic Specificity. Structure *23*, 2087-2098.

Nicoludis, J.M., Vogt, B.E., Green, A.G., Scharfe, C.P.I., Marks, D.S., and Gaudet, R. (2016). Antiparallel protocadherin homodimers use distinct affinity- and specificity-mediating regions in cadherin repeats 1-4. eLife *5*.

Onouchi, T., Kishino-Kaneko, Y., Kameshita, I., Ishida, A., and Sueyoshi, N. (2015). Regulation of Ca2+/calmodulin-dependent protein kinase phosphatase (CaMKP/PPM1F) by protocadherin-γC5 (Pcdh-γC5). Archives of Biochemistry and Biophysics *585*, 109-120.

Pancho, A., Aerts, T., Mitsogiannis, M.D., and Seuntjens, E. (2020). Protocadherins at the Crossroad of Signaling Pathways. Front Mol Neurosci *13*, 117.

Peek, S.L., Mah, K.M., and Weiner, J.A. (2017). Regulation of neural circuit formation by protocadherins. Cellular and Molecular Life Sciences.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth *8*, 785-786.

Rich, R.L., and Myszka, D.G. (2007). Survey of the year 2006 commercial optical biosensor literature. Journal of Molecular Recognition *20*, 300-366.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research *42*, W320-W324.

Rubinstein, R., Goodman, K.M., Maniatis, T., Shapiro, L., and Honig, B. (2017). Structural origins of clustered protocadherin-mediated neuronal barcoding. Seminars in Cell & Developmental Biology *69*, 140-150.

Rubinstein, R., Thu, C.A., Goodman, K.M., Wolcott, H.N., Bahna, F., Mannepalli, S., Ahlsen, G., Chevee, M., Halim, A., Clausen, H*., et al.* (2015). Molecular Logic of Neuronal Self-Recognition through Protocadherin Domain Interactions. Cell *163*, 629-642.

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. Proceedings of the National Academy of Sciences of the United States of America *107*, 14893-14898.

Sergeeva, A.P., Katsamba, P.S., Cosmanescu, F., Brewer, J.J., Ahlsen, G., Mannepalli, S., Shapiro, L., and Honig, B. (2020). DIP/Dpr interactions and the evolutionary design of specificity in protein families. Nat Commun *11*, 2125.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J*., et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology *7*, 539-539.

Steffen, D.M., Ferri, S.L., Marcucci, C.G., Blocklinger, K.L., Molumby, M.J., Abel, T., and Weiner, J.A. (2021). The γ-Protocadherins Interact Physically and Functionally with Neuroligin-2 to Negatively Regulate Inhibitory Synapse Density and Are Required for Normal Social Interaction. Molecular Neurobiology *58*, 2574-2589.

Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006). Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences of the United States of America *103*, 8060-8065.

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.*, et al.* (2014). Structural and energetic determinants of adhesive binding specificity in type I cadherins. Proceedings of the National Academy of Sciences of the United States of America *111*, E4175-E4184.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallographica Section D *67*, 235-242.

Wojtowicz, W.M., Flanagan, J.J., Millard, S.S., and Zipursky, S.L. (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.*, et al.* (2001). Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. Genome Research *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., Honig, B., and Ben-Shaul, A. (2013). Theory and Simulations of Adhesion Receptor Dimerization on Membrane Surfaces. Biophysical Journal *104*, 1221-1229.

Zipursky, S.L., and Grueber, W.B. (2013). The Molecular Basis of Self-Avoidance. In Annual Review of Neuroscience, Vol 36, S.E. Hyman, ed., pp. 547-568.

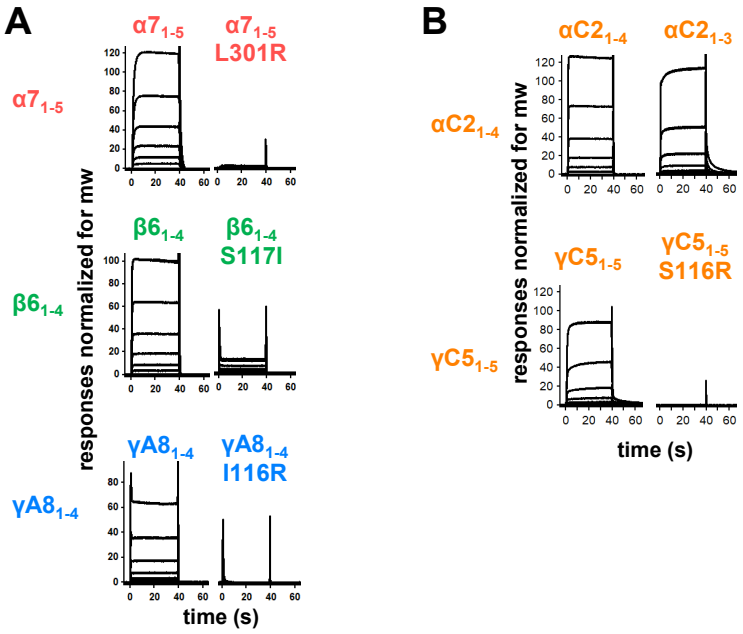


Figure 2—figure supplement 1: *Trans* interface mutants demonstrate homophilic interactions observed in SPR are mediated by the *trans* dimer interface

(A) SPR binding curves for wild-type and *trans* mutant alternate cPcdhs flowed over their respective immobilized wild-type molecule.

(B) SPR binding curves for wild-type and *trans* mutant C-type cPcdhs flowed over their respective immobilized wild-type molecules.

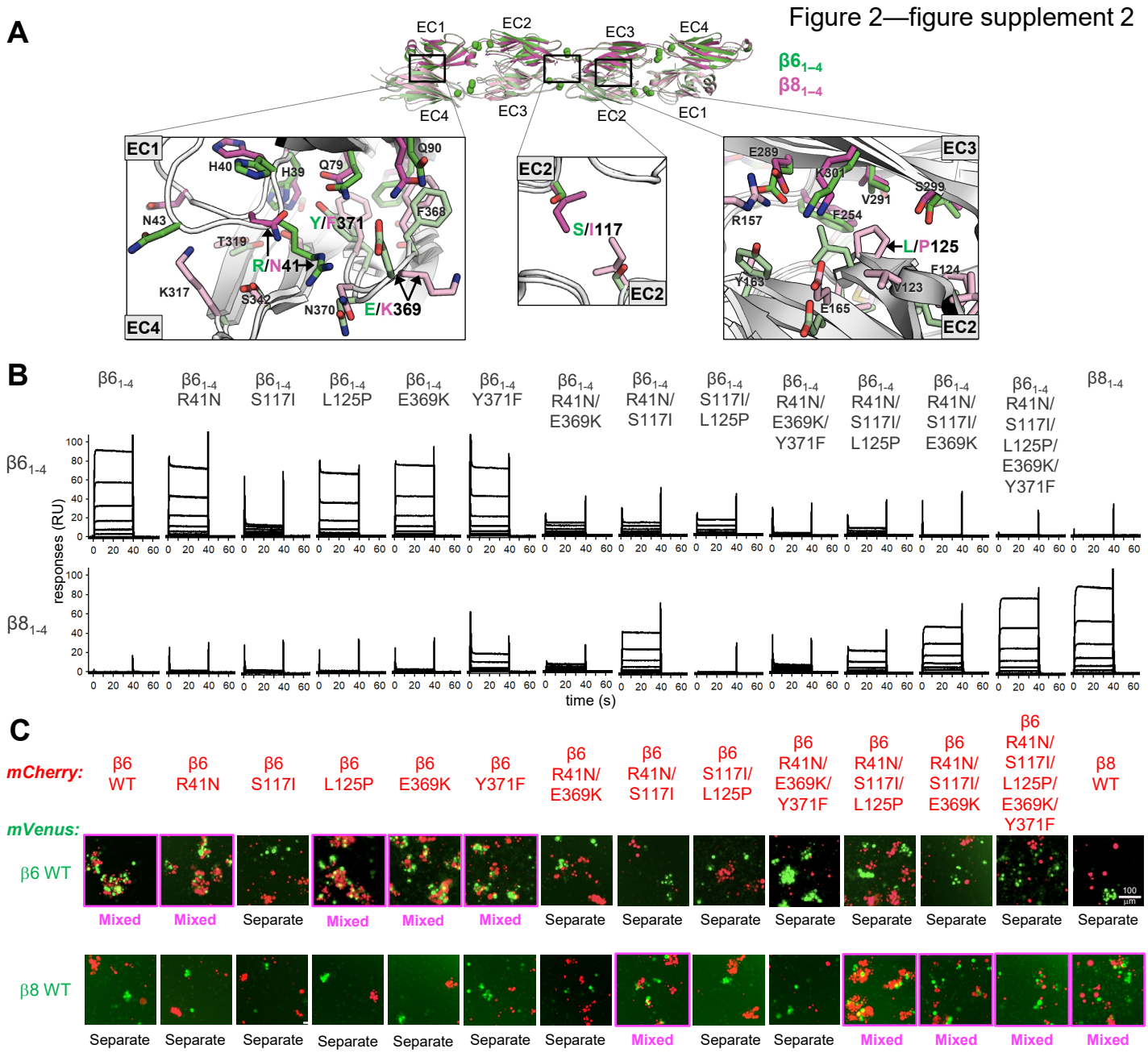


Figure 2 —figure supplement 2: Mutagenesis experiments reveal role in *trans* specificity for the five interfacial residue differences between close pair $\beta6_{1\text{-}4}$ and $\beta8_{1\text{-}4}$

A. Structural superposition of the $\beta6_{1\text{-}4}$ and $\beta8_{1\text{-}4}$ *trans* dimer crystal structures (PDBs: 5DZX and 5DZY) shown in ribbon depiction above, with close-up views of the *trans* interfacial regions containing the five interfacial residues that vary between $\beta6_{1\text{-}4}$ and $\beta8_{1\text{-}4}$ shown below. The two protomers forming the $\beta6_{1\text{-}4}$ dimer are colored green and pale green respectively. The $\beta8_{1\text{-}4}$ dimer is colored magenta/light pink. Bound calcium ions are shown as green spheres. Interfacial residue side chains are shown in the close-up views. The five variable residues are labelled with the $\beta6_{1\text{-}4}$ amino acid given in green and the $\beta8_{1\text{-}4}$ amino acid in magenta: R/N41 is in EC1; E/K369 and Y/F371 are in EC4; S/I117 is in EC2 and self-interacts at the *trans* dimer center of symmetry; and L/P125 is also in EC2.

B. SPR binding profiles of β6 *trans* interface mutants converting $\beta6_{1\text{-}4}$ to $\beta8_{1\text{-}4}$ and the wild-type molecules (shown in columns) were flowed over surfaces coated with wild-type $\beta6_{1\text{-}4}$ or wild-type $\beta8_{1\text{-}4}$ (rows).

C. Results of the K562 co-aggregation assay where cells transfected with mCherry labeled β6 and β8 wild-types (WT) and the same *trans*-specificity mutants as in (**B**) were each mixed with cells transfected with mVenus labeled β6 and β8 wild-types (WT). Experiments where the red and green cells co-aggregate demonstrating interaction between the mCherry-labeled WT or mutant cPcdh and the mVenus-labeled WT cPcdh are labeled "mixed" and highlighted with magenta boxes. Scale bar, 100 µM.

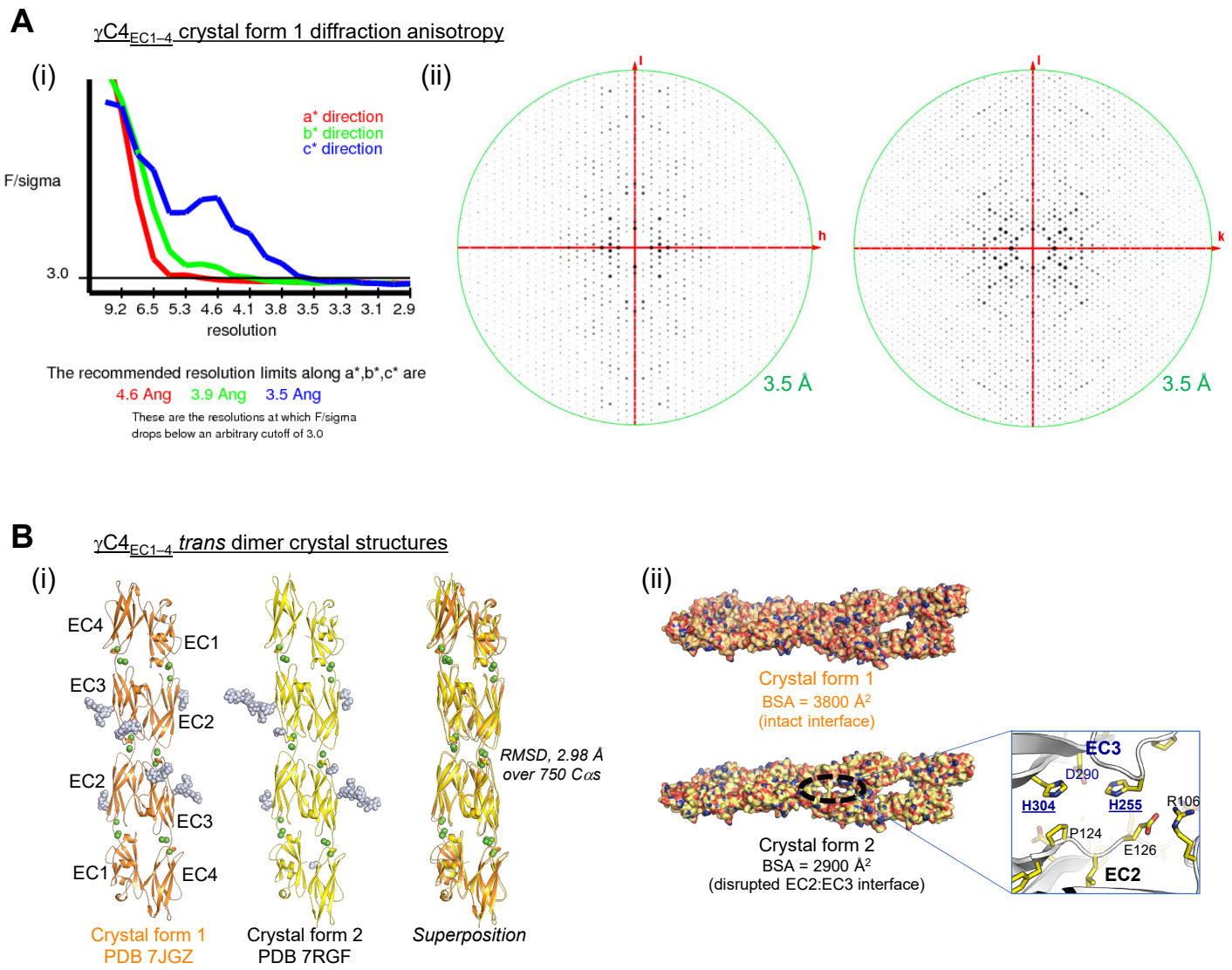


Figure 3—figure supplement 1: γC4 *trans* dimer crystal structures

(A) Our crystallization experiments with $\gamma C4_{EC1-4}$ yielded two distinct crystal forms the first of which showed significant X-ray diffraction anisotropy. (i) UCLA Diffraction Anisotropy Server (Strong et al., 2006) plot shows the F/sigma by resolution along the a*, b* and c* axes. (ii) Synthetic precession photographs of the X-ray diffraction in the k=0 plane (left) and the h=0 plane (right) showing the comparatively stronger/weaker diffraction.

(B) (i) Ribbon diagrams of the $\gamma C4_{EC1-4}$ *trans* dimer crystal structures obtained from the two different crystal forms. Bound calcium ions are shown as green spheres and glycans are shown in pale blue spheres. (ii) The two crystal structures have a markedly different *trans* interface buried surface area (BSA). *Left,* Surface views of the two *trans* dimer crystal structures highlight the difference, with a gap apparent in the EC2:EC3 region of the interface in crystal form 2 that is absent from crystal form 1. *Right,* This disruption in the EC2:EC3 region of the interface observed in the second crystal form could be the impact of the lower pH of the crystallization condition (pH 6) impacting the predominantly electrostatic contacts in this region of the interface including two histidine residues found in the gap region.

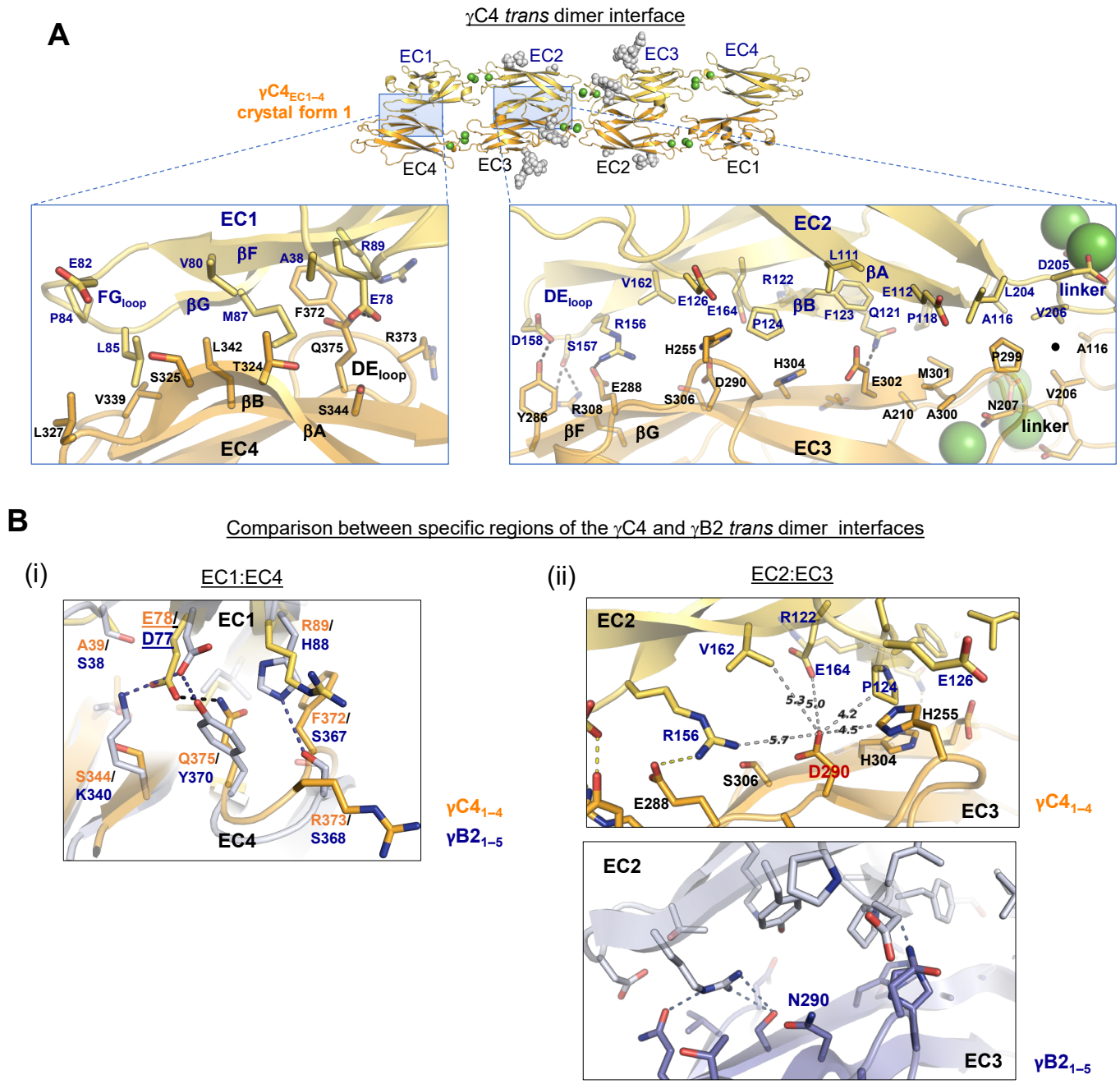


Figure 3—figure supplement 2: Why is the γC4 *trans* dimer interaction so weak?
(**A**) Close up views of the EC1:EC4 and EC2:EC3 interfacial regions from the first crystal form. One protomer in the symmetric dimer is colored orange the other grey. Interfacial residues are labeled, side chains are shown in stick representation and dashed black lines depict potential interfacial hydrogen bond interactions.
(**B**) Comparison between the (i) EC1:EC4 and (ii) EC2:EC3 regions of the γC4 (orange) and γB2 (blue) *trans* dimer interfaces. (i) Structural alignment of the EC1:EC4 portion of the γC4 and γB2 *trans* dimers highlights a possible destabilizing role for γC4 residue E78 since unlike its counterpart in γB2 (D77) it is not juxtaposed with a basic residue. (ii) Similarly, an additional negatively charged residue (D290) which occupies a central position in the γC4 EC2:EC3 interface may also contribute to γC4's comparatively weak *trans* dimer interaction. Distances between the D290 side chain and its nearest contacts are shown as dashed grey lines with distances given in Angstroms.

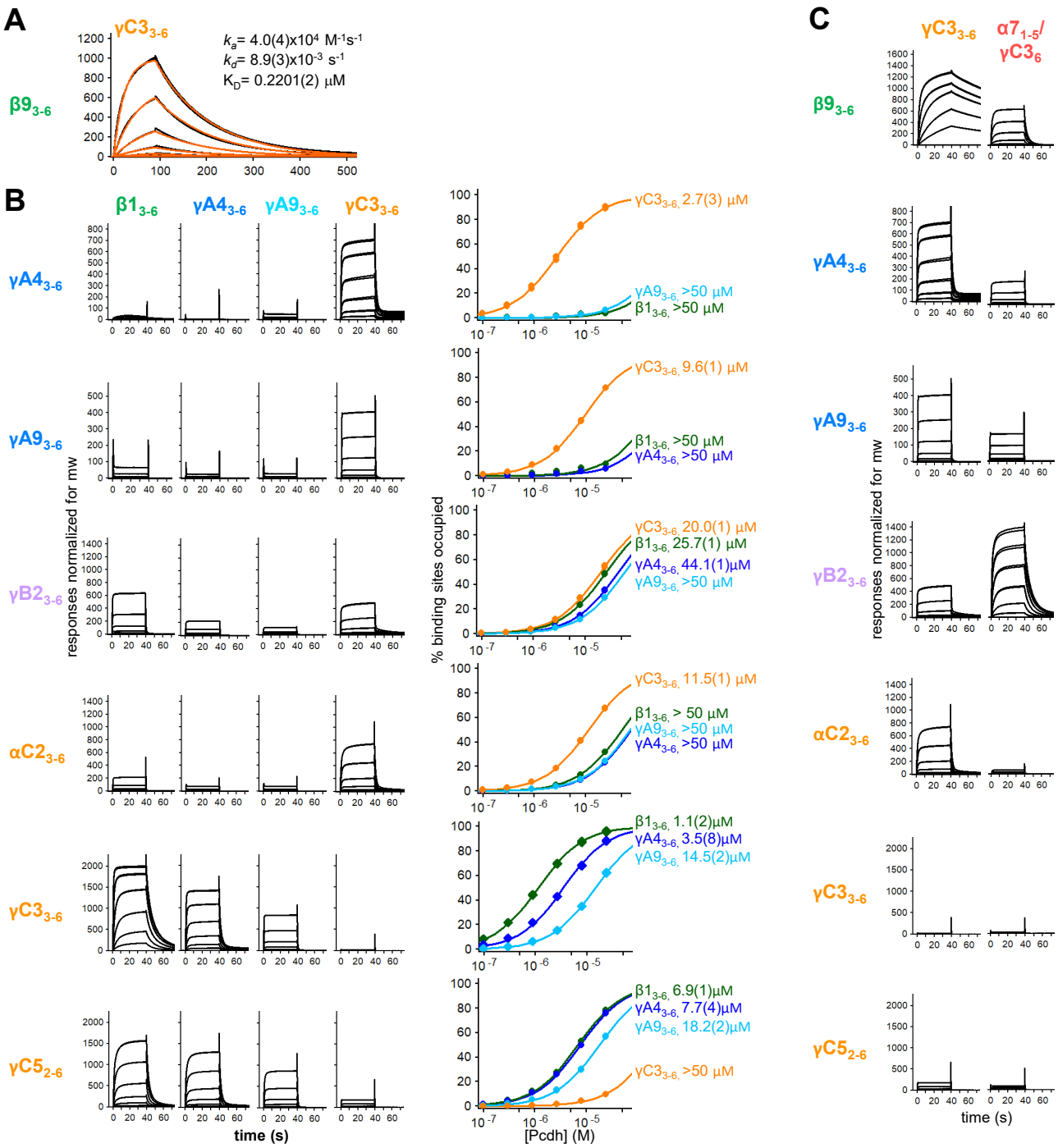


Figure 4—figure supplement 1: Calculation of *cis* interaction dissociation constants and the impact of an α-Pcdh EC5 on family-wide *cis* interactions

(A) Kinetic binding analysis of $\gamma C3_{3\text{-}6}$ analyte binding over a $\beta 9_{3\text{-}6}$ covered surface. Data is shown in black and the red traces represent the fit to an 1:1 binding model.

(B) *Left*, SPR binding profiles from Figure 4 for the four monomeric *cis* fragment analytes over all six *cis* fragment surfaces. *Right*, fit of the binding data for these four analytes to 1:1 binding isotherms to calculate K_Ds. $\gamma A4_{3\text{-}6}$ and $\gamma A9_{3\text{-}6}$ are monomeric and they are not included in the binding isotherms over their respective surface.

(C) SPR binding profiles for $\gamma C3_{3\text{–}6}$ (from Figure 4) and an $\alpha 7_{1\text{–}5}$ / $\gamma C3_6$ chimera flowed over the immobilized *cis* fragment surfaces. Binding profiles for each surface are individually scaled as in Figure 4.

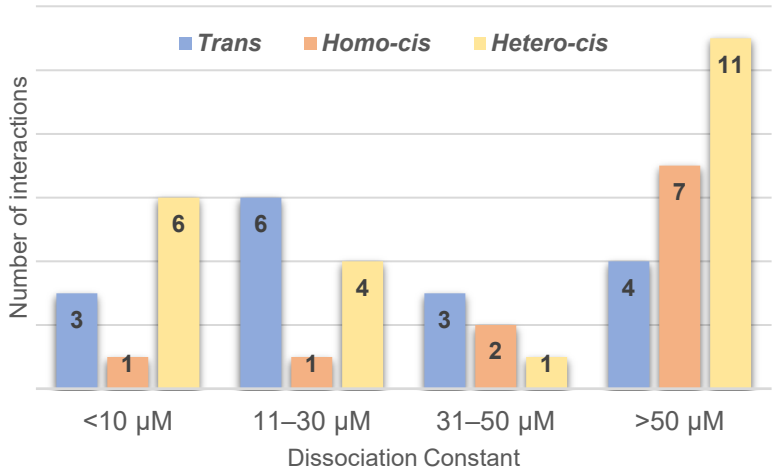


Figure 4—figure supplement 2: cPcdh *cis* and *trans* Dissociation constant, K_Ds
Chart shows the cPcdh *trans* dimer, homophilic *cis* dimer, and heterophilic *cis* dimer interactions for which we have determined binding affinities divided into four subgroups based on their dissociation constant. The *trans* and homophilic *cis* dimer affinities were determined using AUC (Figure 2—source data 1 and Figure 4—source data 1) and the heterophilic *cis* dimer affinities were determined using SPR (Figure 4B). Of the interactions in the >50 μM group one *trans* interaction and four homophilic *cis* interactions are monomeric in solution (> 500 μM K_D). Three of the 10 heterophilic *cis* interactions in the >50 μM group show no binding in our SPR experiments based on a 40 RU binding threshold.

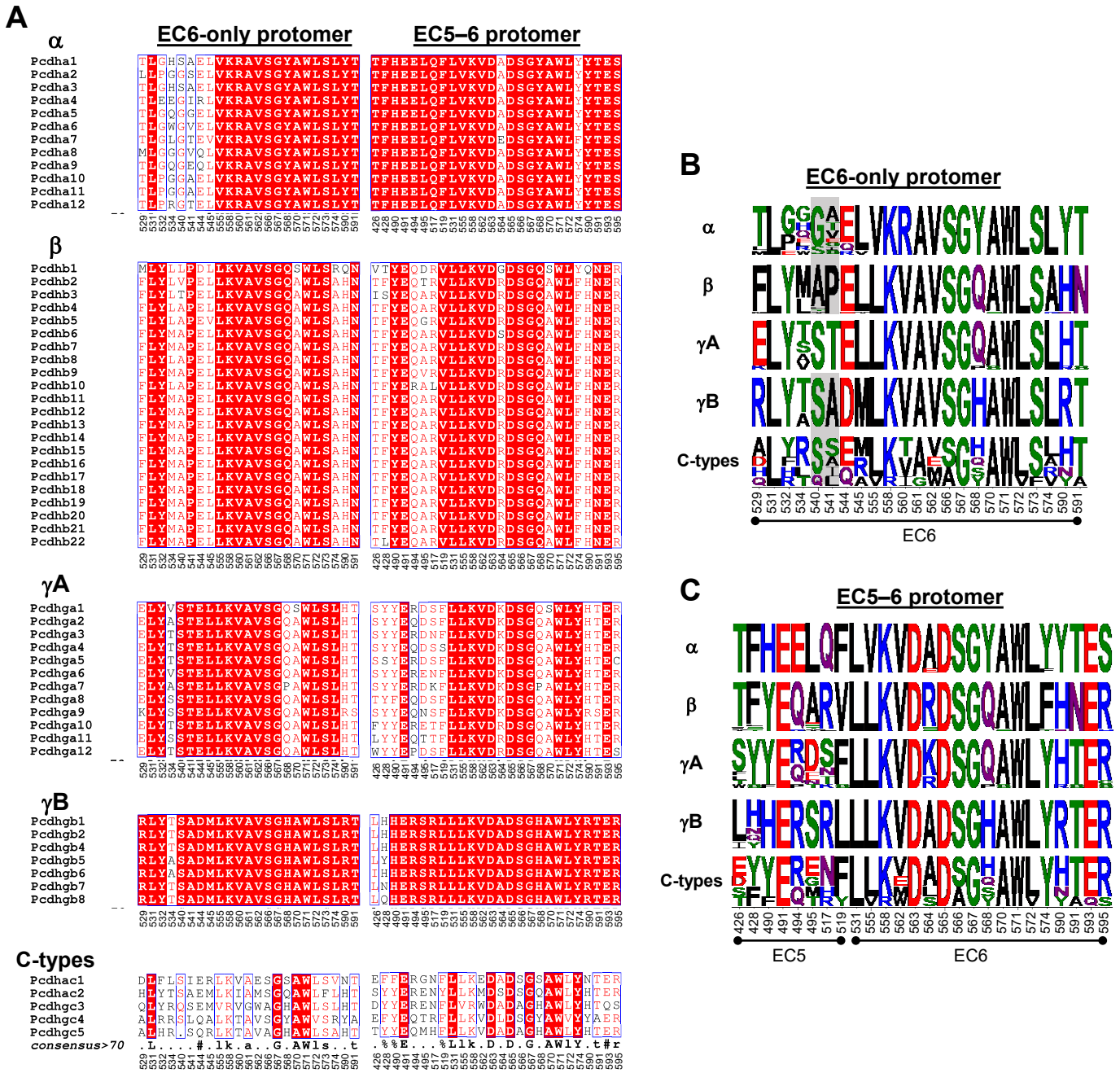


Figure 4—figure supplement 3: Amino acid sequence alignment reveals conservation of *cis* interfacial residues within the alternate cPcdh subfamilies

(A) Amino acid sequence alignments of *cis* interfacial residues from the EC6-only and EC5–6 surfaces for all 58 mouse cPcdhs subdivided by subfamily. Completely conserved residues are highlighted in red with white lettering. Residues 540 and 541 are included in the EC6-only alignments since the crystal structure of γA4 EC3–6 (PDB: 5SZQ) revealed a distinct EC6 A-A'loop architecture to that observed in the γB2, 4, and 7 (PDBs: 5SZR, 6E6B, and 5V5X) *cis* fragment crystal structures that would place these residues in the EC6-only interface if maintained in *cis* interactions.

(B) Sequence logos based on the sequence alignment shown in (**A**) for the EC6-only *cis* interfacial residues from each of the five cPcdh subfamilies highlighting the similarities and conserved differences between the subfamilies. Residues 540 and 541 are included for all isoforms but greyed out for the non-γA isoforms since their involvement may be γA-specific. NB: Previous studies have shown that α-Pcdhs have an impaired EC6-only interface (Thu et al., 2014; Goodman et al., 2017).

(C) Sequence logos for the EC5–6 *cis* interfacial residues from each of the five cPcdh subfamilies

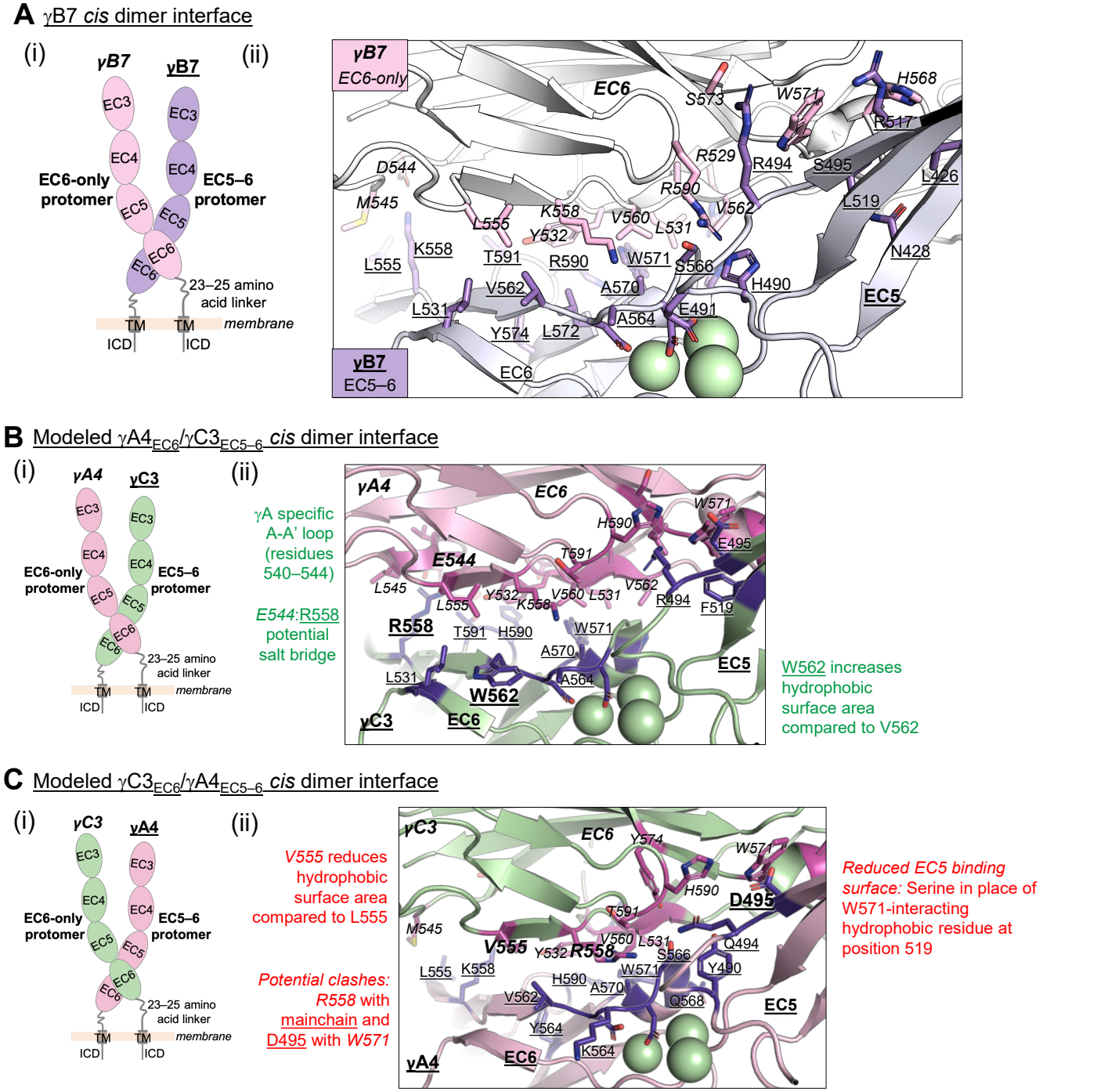


Figure 5—figure supplement 1: Structure-guided sequence analysis of γA4 and γC3 *cis* interactions
(**A**) (i) Schematic of the asymmetric $\gamma B7_{EC3–6}$ *cis* dimer crystal structure. (ii) Close-up view of the γB7 *cis* interface: Interfacial residue side chains are shown in pink for the EC6-only protomer and purple for the EC5–6 protomer. Bound calcium ions are shown as green spheres.
(**B**) (i) Schematic of the $\gamma A4_{EC6}/\gamma C3_{EC5–6}$ *cis* dimer. (ii) Model of the $\gamma A4_{EC6}/\gamma C3_{EC5–6}$ *cis* dimer interaction generated using the γB7 EC3–6 *cis* dimer and monomeric γA4 EC3–6 crystal structures. The model suggests that this will be the preferred orientation for the γA4/γC3 *cis* dimer interaction. Favorable residue differences between γB7 and γA4/γC3 in this orientation are noted in green. The model shown here is only used for hypothesis generation, since it is highly unlikely to be completely accurate.
(**C**) (i) Schematic of the $\gamma C3_{EC6}/\gamma A4_{EC5–6}$ *cis* dimer. (ii) Model of the $\gamma C3_{EC6}/\gamma A4_{EC5–6}$ *cis* dimer. The model suggests that this orientation for the γA4/γC3 *cis* dimer interaction will be disfavored. Unfavorable residue differences between γB7 and γA4/γC3 in this orientation are noted in red.

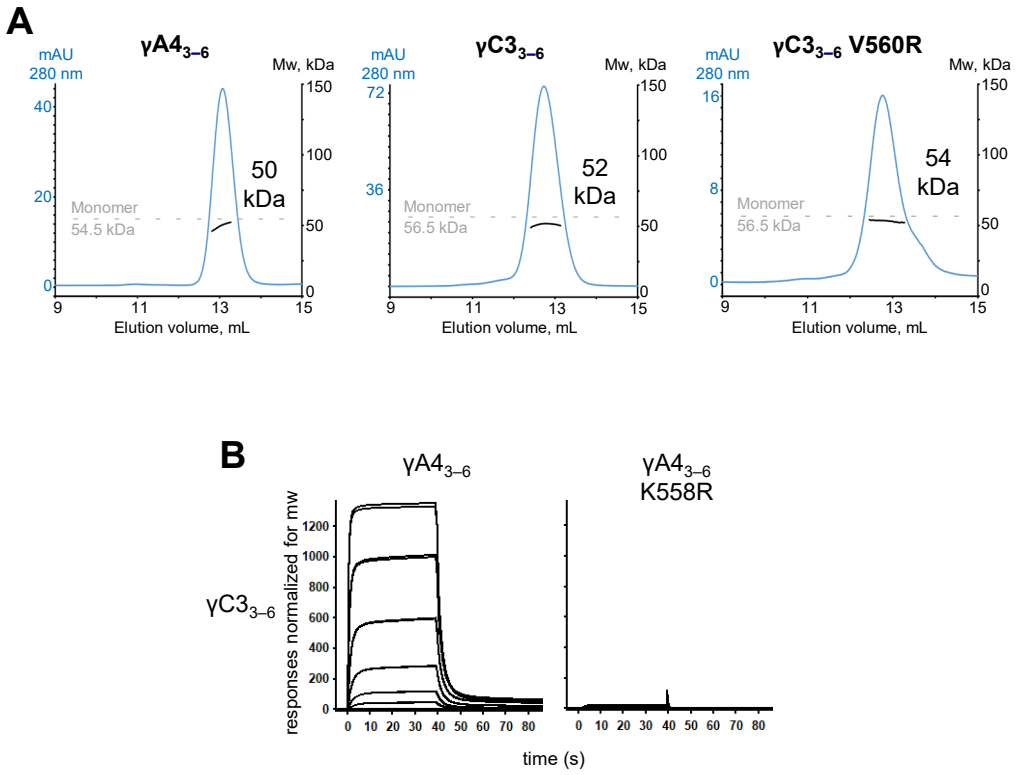


Figure 5—figure supplement 2: γA4 and γC3 *cis*-fragments behave as monomers in SEC-MALS and mutating γA4 to make it more like γC3 prevents γA4/γC3 *cis*-heterodimerization

(**A**) SEC-MALS data for wild-type $\gamma A4_{3-6}$, wild-type $\gamma C3_{3-6}$, and $\gamma C3_{3-6}$ V560R showing all three molecules are monomeric in SEC-MALS, consistent with their behavior in sedimentation equilibrium AUC. Plots show size exclusion absorbance at 280 nm trace in blue (left axis), molecular weight of the eluant peak in black (right axis), and the monomer molecular weight of $\gamma A4_{3-6}$ or $\gamma C3_{3-6}$ measured by mass spectrometry – 54.5 kDa and 56.5 kDa respectively – as dashed grey lines. Average molecular weight of the molecules in the eluant peaks are labeled.
(B) SPR binding profiles for $\gamma A4_{3-6}$ wild type and $\gamma A4_{3-6}$ with γC3-like *cis* interface mutation K558R flowed over immobilized wild-type $\gamma C3_{3-6}$. Loss of $\gamma C3_{3-6}$ interaction in the presence of the K558R mutation is consistent with the SEC-MALS results shown in Figure 5.

Protein	Oligomeric State	Dissociation Constant, K_D (μM)
Trans-interacting fragments		
$\alpha4_{1-5}$	Dimer	5.0 ± 0.80
$\alpha7^*_{1-5}$	Dimer	2.91 ± 0.55
$\alpha12_{1-5}$	Dimer	34 ± 2.8
$\beta6^*_{1-4}$	Dimer	16.3 ± 2.1
$\beta8^*_{1-4}$	Dimer	24.0 ± 0.43
$\gamma A1^*_{1-4}$	Dimer	13.3 ± 0.93
$\gamma A4^*_{1-4}$	Dimer	45.3 ± 1.52
$\gamma A8^*_{1-4}$	Dimer	30 ± 1.5
$\gamma A9^*_{1-5}$	Dimer	8.61 ± 0.35
$\gamma B2^*_{1-5}$	Dimer	21.8 ± 0.21
$\gamma B4_{1-5}$	Dimer	38 ± 0.33
$\gamma B5^*_{1-4}$	Dimer	79.1 ± 4.3
$\gamma B5_{1-4}$-AVI	Dimer	50 ± 0.4
$\alpha C2^*_{1-4}$	Dimer	20.6 ± 1.19
$\gamma C3^*_{1-4}$	Dimer	115 ± 1.49 (K_i/K_D = 1.57)
$\gamma C4_{1-4}$	Very weak dimer	> 500†
$\gamma C5^*_{1-5}$	Dimer	100 ± 4.33
Trans mutants		
$\alpha7_{1-5}$ L301R	Weakly dimeric	490 ± 57
$\gamma A8_{1-4}$ I116R*	Monomer	N/A
$\alpha C2^*_{1-3}$	Dimer	242 ± 0.1 (K_i/K_D = 1.48)
β_{1-4} R41N	Dimer	160 ± 0.38
β_{1-4} S117I	Dimer	72 ± 34
β_{1-4} L125P	Dimer	150 ± 20
β_{1-4} E369K	Dimer	23 ± 2.8
β_{1-4} Y371F	Dimer	39 ± 5.6
β_{1-4} R41N/S117I	Precipitate	N/A
β_{1-4} R41N/E369K	Dimer	41 ± 0.69
β_{1-4} S117I/L125P	Dimer	68 ± 5.1
β_{1-4} R41N/S117I/L125P	Weak Dimer	350 ± 11
β_{1-4} R41N/S117I/E369K	Dimer	32 ± 0.95
β_{1-4} R41N/S117I/Y371F	Dimer	18 ± 0.11
β_{1-4} R41N/S117I/L125P/E369K/Y371F	Dimer	63 ± 11

Figure 2—source data 1. Sedimentation equilibrium analytical ultracentrifugation data for *trans* SPR reagents

* Previously published data (Rubinstein et al., 2015; Goodman et al., 2016a; Goodman et al., 2016b)
† Dissociation constants larger than 500 μM cannot be accurately determined.

	cPcdh γC4$_{EC1–4}$ crystal form 1		cPcdh γC4$_{EC1–4}$ crystal form 2
Data collection			
Date	02/07/2017		03/04/2017
Beamline	APS 24-ID-E		APS 24-ID-E
Wavelength (Å)	0.97918		0.97918
Space group	$I2_12_12_1$		$P2_12_12_1$
Cell dimensions			
a, b, c (Å)	75.220, 115.680, 194.480		51.944, 103.948, 200.592
α, β, γ (°)	90, 90, 90		90, 90, 90
	Spherical resolution limits	*Ellipsoidal resolution limits*	
Resolution (Å)	40.00–3.50 (3.83–3.50)	40–4.6/3.9/3.5 (3.84–3.51)	40.00–2.40 (2.49–2.40)
No. of reflections	79877 (19206)	53652 (2402)	240907 (16849)
Unique reflections	11071 (2608)	7545 (340)	42979 (4175)
R_{merge}	0.381 (3.459)	0.215 (0.830)	0.060 (0.438)
R_{meas}	0.411 (3.722)	0.232 (0.896)	0.073 (0.560)
R_{pim}	0.152 (1.366)	0.086 (0.335)	0.040 (0.342)
CC(1/2)	0.995 (0.533)	0.997 (0.829)	0.999 (0.924)
I/σI	5.7 (0.8)	8.3 (2.9)	13.3 (2.5)
Spherical completeness (%)	99.9 (100.0)	67.6 (12.5)	98.9 (92.6)
Ellipsoidal completeness (%)	N/A	96.8	N/A
Redundancy	7.2 (7.4)	7.1 (7.1)	5.6 (4.0)
Refinement			
Resolution (Å)	40–4.6/3.9/3.5		40.00–2.40
Unique reflections	7545		42815
Completeness in diffracting sphere or ellipsoid* (%)	96.8*		98.5
R_{work} / R_{free} (%)	22.3 / 27.1		19.6 / 24.1
Molecules in A.S.U.	1		2
Number of atoms	*3242*		*6736*
Protein	3120		6327
Ligand/Ion	122		284
Water	0		125
B-factors	*101.24*		*73.39*
Protein	101.03		72.99
Ligand/Ion	106.73		88.54
Water	N/A		59.26
R.m.s. deviations			
Bond lengths (Å)	0.004		0.003
Bond angles (°)	0.889		0.653
Ramachandran			
Favored (%)	98.80		98.92
Allowed (%)	1.20		1.08
Outliers (%)	0.00		0.00
Rotamer outliers (%)	0.29		0.57
Wilson B	77.57		51.79
PDB ID	7JGZ		7RGF

Figure 3—source data 1. X-ray crystallography data collection and refinement statistics

Values in parentheses are for the outer shell. APS, Advanced Photon Source, Argonne National Lab; A.S.U., asymmetric unit; R.m.s., Root mean square. See Figure 4—figure supplement 1 and Methods for further details on the ellipsoidal resolution limits.

RMSDs between EC1–4 dimers	Mouse cPcdh γC4 dimer (crystal form 1)
Alternate clustered Pcdhs:	
Mouse cPcdh α7 dimer	3.3 Å (801 atoms)
Mouse cPcdh β6 dimer	4.5 Å (769 atoms)
Mouse cPcdh γA1 dimer	3.5 Å (794 atoms)
Mouse cPcdh γB2 dimer	4.3 Å (802 atoms)
Non-clustered δ1 Pcdhs:	
Human ncPcdh1 dimer	7.0 Å (784 atoms)
Non-clustered δ2 Pcdhs:	
Xenopus ncPcdh8.1 dimer	3.4 Å (756 atoms)
Human ncPcdh10 dimer	2.4 Å (785 atoms)
Human ncPcdh18 dimer	3.1 Å (786 atoms)
Human ncPcdh19 dimer	3.3 Å (780 atoms)
Zebrafish ncPcdh19 dimer	2.4 Å (778 atoms)

Figure 3—source data 2. Overall structural similarity between γC4, alternate cPcdhs, and non-clustered Pcdhs *trans* dimer structures

Table lists the pairwise root mean square deviations over aligned EC1–EC4 Cα's (RMSDs) between the intact $\gamma C4_{EC1–4}$ *trans* dimer (crystal form 1) and a representative selection of available cPcdh and ncPcdh *trans* dimer structures. The number of aligned Cα's is given in parentheses. The PDB codes for the aligned crystal structures are as follows: $\alpha7_{EC1–5}$, 5DZV; $\beta6_{EC1–4}$, 5DZX; $\beta8_{EC1–4}$, 5SZL; $\gamma B2_{EC1–5}$, 5T9T; $ncPcdh1_{EC1–4}$, 6MGA; $ncPcdh8.1_{EC1–6}$, 6VG1; $ncPcdh10_{EC1–4}$, 6VFW; $ncPcdh18_{EC1–4}$, 6VFR; $ncPcdh19_{EC1–4}$, 6VFU; and zebrafish $ncPcdh19_{EC1–4}$, 5IU9 (Goodman et al., 2016a; Goodman et al., 2016b; Modak and Sotomayor, 2019; Harrison et al., 2020; Cooper et al., 2016).

RMSDs over individual chains	$\gamma C4_{EC1–4}$ crystal form 1
$\gamma C4_{EC1–4}$ crystal form 2 chain A	2.8 Å (364 atoms)
$\gamma C4_{EC1–4}$ crystal form 2 chain B	2.4 Å (370 atoms)
$\alpha 7_{EC1–5}$	2.2 Å (364 atoms)
$\beta 6_{EC1–4}$	2.5 Å (371 atoms)
$\gamma A1_{EC1–4}$	2.9 Å (390 atoms)
$\gamma B2_{EC1–5}$	2.8 Å (398 atoms)
$\alpha C2_{EC1–3}$	2.3 Å (295 atoms)
$\gamma C3_{EC1–3}$	2.2 Å (282 atoms)
$\gamma C5_{EC1–3}$	2.2 Å (308 atoms)
Human ncPcdh1$_{EC1–4}$	4.1 Å (360 atoms)
Xenopus ncPcdh8.1$_{EC1–6}$	1.9 Å (337 atoms)
Human ncPcdh10$_{EC1–4}$	1.9 Å (362 atoms)
Human ncPcdh18$_{EC1–4}$	1.9 Å (372 atoms)
Human ncPcdh19 $_{EC1–4}$	3.1 Å (385 atoms)
Zebrafish ncPcdh19$_{EC1–4}$	2.3 Å (391 atoms)

Figure 3—source data 3. Structural similarity between individual protomers of γC4, alternate cPcdhs, and non-clustered Pcdhs

Table lists the pairwise root mean square deviations over aligned Cα's (RMSDs) between $\gamma C4_{EC1–4}$ and individual protomers from a representative selection of available cPcdh and ncPcdh structures. Chain A was used for all alignments except where noted. The number of aligned Cα's for each pairwise alignment are given in parentheses. The PDB codes for the aligned crystal structures are as follows: $\alpha 7_{EC1–5}$, 5DZV; $\beta 6_{EC1–4}$, 5DZX; $\beta 8_{EC1–4}$, 5SZL; $\gamma B2_{EC1–5}$, 5T9T; $\alpha C2_{EC1–3}$, 4ZPM; $\gamma C3_{EC1–3}$, 4ZI8; $\gamma C5_{EC1–3}$, 4ZPO; ncPcdh1$_{EC1–4}$, 6MGA; ncPcdh8.1$_{EC1–6}$, 6VG1; ncPcdh10$_{EC1–4}$, 6VFW; ncPcdh18$_{EC1–4}$, 6VFR; ncPcdh19$_{EC1–4}$, 6VFU; and zebrafish ncPcdh19$_{EC1–4}$, 5IU9 (Goodman et al., 2016a; Goodman et al., 2016b; Rubinstein et al., 2015; Nicoludis et al., 2015; Modak and Sotomayor, 2019; Harrison et al., 2020; Cooper et al., 2016).

RMSDs over interacting domains in *trans* dimers	γC4 crystal form 1 (intact interface)	
	EC1:EC4 (196 atoms)	EC2:EC3 (209 atoms)
γC4 crystal form 2 chain A:chain B	1.3 Å (196 atoms)	0.9 Å (198 atoms)
γC4 crystal form 2 chain B:chain A	1.9 Å (194 atoms)	1.3 Å (207 atoms)
α7 dimer	2.0 Å (189 atoms)	1.7 Å (196 atoms)
β6 dimer	2.5 Å (180 atoms)	3.8 Å (200 atoms)
γA1 dimer	3.2 Å (191 atoms)	2.2 Å (182 atoms)
γB2 dimer	1.5 Å (187 atoms)	1.8 Å (196 atoms)
Human ncPcdh1 dimer	1.5 Å (139 atoms)	2.9 Å (185 atoms)
Xenopus ncPcdh8.1 dimer	3.0 Å (174 atoms)	2.4 Å (206 atoms)
Human ncPcdh10 dimer	2.1 Å (189 atoms)	1.3 Å (195 atoms)
Human ncPcdh18 dimer	1.7 Å (157 atoms)	2.4 Å (207 atoms)
Human ncPcdh19 dimer	2.0 Å (155 atoms)	1.2 Å (187 atoms)
Zebrafish ncPcdh19 dimer	1.9 Å (165 atoms)	1.2 Å (198 atoms)

Figure 3—source data 4. Overall structural similarity between EC1:EC4 and EC2:EC3 interface regions in the γC4 *trans* dimer and *trans* dimers from each alternate cPcdh subfamily, a δ1 non-clustered Pcdh, and five δ2 non-clustered Pcdhs

Root mean square deviations over aligned Cα's (RMSDs) between pairs of interacting EC1:EC4 and EC2:EC3 regions of Pcdh *trans*-dimer structures are shown. The number of aligned Cα's for each pair is given in parentheses. The PDB codes for the aligned crystal structures are as follows: $\alpha7_{EC1–5}$, 5DZV; $\beta6_{EC1–4}$, 5DZX; $\beta8_{EC1–4}$, 5SZL; $\gamma B2_{EC1–5}$, 5T9T; $\alpha C2_{EC1–3}$, 4ZPM; $\gamma C3_{EC1–3}$, 4ZI8; $\gamma C5_{EC1–3}$, 4ZPO; $ncPcdh1_{EC1–4}$, 6MGA; $ncPcdh8.1_{EC1–6}$, 6VG1; $ncPcdh10_{EC1–4}$, 6VFW; $ncPcdh18_{EC1–4}$, 6VFR; $ncPcdh19_{EC1–4}$, 6VFU; and zebrafish $ncPcdh19_{EC1–4}$, 5IU9 (Goodman et al., 2016a; Goodman et al., 2016b; Rubinstein et al., 2015; Nicoludis et al., 2015; Modak and Sotomayor, 2019; Harrison et al., 2020; Cooper et al., 2016)

	Sequence identity over EC1–EC4 (%)				
	αC1	αC2	γC3	γC4	γC5
C-type cPcdhs					
αC1	100	46	45	46	45
αC2	46	100	49	49	52
γC3	45	49	100	49	46
γC4	46	49	49	100	54
γC5	45	52	46	54	100
δ1 non-clustered Pcdhs					
pcdh1	35	38	37	36	36
pcdh7	37	37	38	34	38
pcdh9	35	34	37	33	34
pcdh11x	35	35	35	35	34
δ2 non-clustered Pcdhs					
pcdh8	36	42	40	42	40
pcdh10	45	52	53	49	52
pcdh17	44	49	46	45	44
pcdh18	34	40	43	39	40
pcdh19	40	44	43	43	44
Zebrafish pcdh19	39	45	43	41	42
Other non-clustered Pcdhs					
pcdh12	35	41	40	39	39
pcdh16	32	35	36	37	31
pcdh20	30	33	31	29	31
Alternate cPcdhs					
α4	43	47	48	41	44
α7	44	48	48	45	46
α12	43	48	49	44	44
β6	41	42	44	41	40
β8	40	43	45	42	40
γA4	46	46	48	46	46
γA8	46	46	47	44	49
γA9	47	46	47	44	49
γB2	40	42	49	44	44
γB4	41	43	45	43	45
γB5	42	43	47	44	45

Figure 3—source data 5. Sequence identity over *trans*-interacting EC1–4 regions of C-type cPcdhs with C-type cPcdhs, non-clustered Pcdhs, and alternate cPcdhs

Protein	Oligomeric State	Dissociation Constant, K_D (μM)
Cis-interacting fragments		
$\beta1_{3-6}$	Monomer	N/A
$\beta6_{1-6}$	Tetramer	1.7 / 12.1†
$\beta9_{3-6}$	Dimer	35 ± 3.1
$\gamma A3_{3-6}$	Dimer	110 ± 7.3
$\gamma A4^*_{3-6}$	Monomer	N/A
$\gamma A9_{3-6}$	Monomer	N/A
$\gamma B2^*_{3-6}$	Dimer	80.1 ± 12.8
$\gamma B5^*_{3-6}$	Dimer	32.6 ± 4.6
$\gamma B7^*_{3-6}$	Dimer	59.0 ± 3.4
$\alpha C2_{3-6}$-AVI	Dimer	7.2 ± 1.2
$\alpha C2^*_{2-6}$	Dimer	8.92 ± 0.28
$\alpha7_{1-5}/\gamma C3_6$ chimera*	Tetramer	3.0 / 3.9†
$\gamma C3^*_{3-6}$	Monomer	N/A
$\gamma C5_{2-6}$	Dimer	18.4 ± 0.24
Cis mutants		
$\gamma B7_{3-6}$ Y532G*	Monomer	N/A
$\gamma B7_{3-6}$ A570R	Monomer	N/A

Figure 4—source data 1. Sedimentation equilibrium analytical ultracentrifugation data for *cis* SPR reagents

* Previously published data (Rubinstein et al., 2015; Goodman et al., 2016b; Goodman et al., 2017)
† K_Ds of monomer-to-dimer / dimer-to-tetramer transitions from fitting the data to a tetramer model.