Supplemental Experimental Procedures

Protein production

Pcdh cDNA for Pcdhα7 EC1-3 (Q1–D316) and EC1-5 (Q1–D530); PcdhαC2 EC1-3 (Q1–D318), EC1-4 (Q1–D422), EC1-5 (Q1–D532), EC1-6 (Q1–K644) and EC2-6 (S103–K644); Pcdhβ1 EC1-3 (A1–D314); PcdhγA8 EC1-3 (Q1–D313), EC1-4 (Q1–D418) and EC2-6 (N101–E641); PcdhγB6 EC1-4 (G1–D418), EC1-5 (G1–D528) and EC1-6 (G1–E641); PcdhγC5 EC1-3 (Q1–D316), EC1-3 with extended N-terminus (including GWSSG before Q1), EC1-4 (Q1–D420), EC1-5 (Q1–D530), EC1-6 (Q1–E643) and EC2-6 (S101–E643), excluding the predicted signal sequence, were cloned into a modified pαSHP-H mammalian expression vector (a kind gift from Daniel J. Leahy, John Hopkins University) with a BiP signal sequence and a C-terminal octahistidine tag. The cDNA for PcdhγC5 EC1-3 was codon optimized to improve protein yields. The γA8 I116R, αC2 S118R and γC5 S116R mutations were introduced by the Quikchange method (Stratagene). These constructs were transfected using polyethyleneimine (Polysciences Inc.) into suspension-adapted HEK293 Freestyle cells (Invitrogen) in serum free media (Invitrogen). The media was harvested 6 days after transfection and the secreted proteins were purified by nickel affinity chromatography followed by size exclusion chromatography. Purified proteins were concentrated to 2-23mg/ml for AUC or crystallography experiments.

Crystallography

Pcdh crystals were grown by vapor diffusion in 1-2μl hanging drops, except the Pcdhβ1 EC1-3 crystals, which were grown in 0.2μl sitting drops. The crystallization conditions were: 28% PEG MME 500, 100mM sodium acetate, pH 4 for PcdhαC2 EC1-3; 24% (w/v) PEG1500, 20% (v/v) glycerol, 3% (w/v) glucose for Pcdhβ1 EC1-3; 28% PEG400, 100mM Tris-Cl, pH 8.7 for

PcdhγA8 EC1-3; 40% (v/v) MPD, 5% (w/v) PEG 8000, 100mM sodium cacodylate, pH 6 for the PcdhγC5 EC1-3 $P4_32_12$ crystal form; 16% (w/v) PEG 6000, 200mM calcium acetate, 100mM imidazole, pH 8.0 (30% (v/v) PEG 400 added cryoprotectant) for the PcdhγC5 EC1-3 C2 crystal form; 25.5% (w/v) PEG 4000, 15% (v/v) glycerol, 3mM calcium chloride, 85mM Tris-Cl, pH 8.5 (20% (v/v) ethylene glycol added cryoprotectant) for the PcdhγC5 EC1-3 $P2_1$ crystal form; 8% (w/v) PEG3350, 200mM potassium nitrate, 3% (v/v) glycerol (30% PEG400 added cryoprotectant) for PcdhγC5 EC1-3 with extended N-terminus. Heavy atom derivatives of the PcdhγC5 EC1-3 $P4_32_12$ crystals were obtained by soaking the crystals in the crystallization condition supplemented with 1mM ethyl mercuric phosphate (EMP) or K_2HgI_4 for 2-16h.

Complete native and derivative datasets were collected from single crystals at 100K on either the Northeastern Collaborative Access Team beamline 24-ID-E at the Advanced Photon Source, Argonne National Laboratory or beamline X4C at National Synchrotron Light Source, Brookhaven National Laboratory. The PcdhγC5 EC1-3 crystal data was indexed using DENZO and scaled and merged with SCALEPACK (Otwinowski and Minor, 1997, 2001). All other data was indexed with MOSFLM (Battye et al., 2011) and scaled and merged with Scala (PcdhαC2 EC1-3 and PcdhγC5 EC1-3 extended N-terminus) or Aimless (PcdhγA8 EC1-3 and Pcdhβ1 EC1-3) (Evans, 2007).

The PcdhγC5 EC1-3 $P4_32_12$ crystal structure was solved using the MIRAS technique. Initial heavy atom sites were located using SOLVE/RESOLVE (Terwilliger and Berendzen, 1999). Optimization of heavy atom sites and solvent flattening was then carried out in SHARP (delaFortelle and Bricogne, 1997) to generate the initial electron density map. Initial model building into the map was carried out in Coot (Emsley et al., 2010) and iterative refinement and model building were carried out using Phenix (Adams et al., 2010) and Coot.

All other Pcdh crystal structures were solved by molecular replacement with Phaser (McCoy et al., 2007) using the PcdhγC5 EC1-3 $P4_32_12$ crystal structure as a search model, except Pcdhβ1 EC1-3, for which PcdhαC2 EC1-3 was used as the search model. Iterative refinement and model building were then conducted using Phenix and Coot.

Cell aggregation assay

Plasmids. DNA fragments encoding fluorescent fusion full length Pcdh isoforms were generated as previously described (Thu et al., 2014). The domain deletion and domain swapping between different Pcdh isoforms were made by performing overlapped PCR. The arginine mutations, the double mutations, and the mutations between close pairs were generated by the Quikchange method (Stratagene).The PCR products were then sub-cloned into gateway entry vectors and corresponding expression vectors. EC domains were assigned as previously described (Thu et al., 2014). Transmembrane domains (TM) were predicted by using TMHIM web (http://www.cbs.dtu.dk/services/TMHMM-2.0/). Signal peptides (SP) were predicted by SignalP prediction tools from SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/). Primer sequences used for PCR amplifications in domain deletion/swapping studies and site directed mutagenesis will be provided upon request.

Cell aggregation assays. Aggregation assays were performed as previously described (Thu et al., 2014). Expression constructs generated by gateway cloning system were transfected into K562 cells (human leukemia cell line, ATCC CCL243) by electroporation method using an Amaxa 4D-Nucleofactor (Lonza). The transfected cells were grown in culture for 24 hours. Cells were then allowed to aggregate for one to three hours on a rocker inside an incubator at 37ºC.

The cells were then fixed in 4% PFA for 10 minutes, washed in PBS, and cleared with 50% glycerol for imaging. The images were taken using an Olympus fluorescent microscope.

Co-aggregation assays. Differentially tagged wild-type or modified Pcdh expression constructs were transfected into K562 cells as described above. K562 cells expressing mCherry- or mVenus-tagged Pcdhs were mixed after 24 hours by shaking for one to three hours. Images of red and green cell aggregates were taken using an Olympus fluorescent microscope.

Effect of the Ca^{2+} chelators EDTA or EGTA on cell aggregation. To assess the requirement for calcium ions in cell-cell adhesion mediated by over-expressed N-cadherin and four Pcdh isoforms, K562 cells transfected with constructs encoding these proteins were treated with EDTA (10mM) or EGTA (5mM) prior to performing the cell aggregation assay. The assay was performed as described above and the fluorescent images were taken with an Olympus inverted microscope.

Sedimentation equilibrium Analytical Ultracentrifugation

Experiments were performed in a Beckman XL-A/I analytical ultracentrifuge (Beckman-Coulter, Palo Alto CA, USA), utilizing six-cell centerpieces with straight walls, 12 mm path length, and sapphire windows. Samples were dialyzed over-night and then diluted in 10 mM Tris, 150 mM NaCl, 3 mM $CaCl_2$, pH 8.0, with varying concentrations of imidazole, as follows: 100 mM (Pcdhα7 EC1-3 and EC1-5; Pcdhβ1 EC1-3; PcdhγC5 EC1-3, EC1-3 extended N-term, and EC1-5; PcdhγA8 EC1-3 and EC2-6; PcdhαC2 EC1-5; PcdhγC5 EC1-4 and EC1-5) 200 mM (PcdhαC2 EC1-3 EC1-4, EC1-6, EC2-6, and EC1-6 S118R; PcdhγA8 EC1-4 and EC1-4 I116R; PcdhγC5 EC2-6 and EC2-6 S116R) or 250 mM (PcdhγC5 EC1-6 and EC1-6 S118R; PcdhγB6 EC1-6). Proteins were diluted to an absorbance at 10 mm and 280 nm of 0.65, 0.43 and 0.23 in

channels A, B and C, respectively. Dilution buffer was used as blank. All samples were run at four speeds, the lowest speed held for 20h then four scans with 1h interval, the second lowest held for 10h then four scans with 1h interval, the third lowest and the highest speed as the second lowest. The speeds were 11000, 14000, 17000 and 20000 rpm (all EC 1-3 constructs) or 9000, 11000, 13000 and 15000 rpm (all EC 1-4, EC1-5 and EC 1-6 constructs), respectively. Measurements were done at 25^{o}C, and detection was by UV at 280 nm. Solvent density and protein v-bar were determined using the program SednTerp. (Alliance Protein Laboratories, Corte Cancion, Thousand Oaks, CA, USA) For calculation of dimeric K_D and apparent molecular weight, all useful data were used in a global fit, using the program HeteroAnalysis, obtained from University of Connecticut. (www.biotech.uconn.edu/auf) Calculation of tetrameric K_D was carried out with the program Sedphat (http://www.analyticalultracentrifugation.com/sedphat/index.htm).

O-mannosylation

Mass spectrometric analyses were performed essentially as previously described (Halim et al., 2015). Briefly, 5 µg of each protein was reduced (5 mM dithiothreitol, 60 °C, 30 min) and alkylated (10 mM iodoacetamide, RT, 30 min) before a 16h, 37°C incubation with 1 µg trypsin (Roche). Tryptic digests were analyzed on a setup composed of an EASY-nLC 1000 UHPLC (Thermo Scientific) interfaced via a nanoSpray Flex ion source to an LTQ-Orbitrap Velos Pro hybrid mass spectrometer. The analytical column (PicoFrit Emitters, New Objectives, 75 µm inner diameter) was packed in-house with Reprosil-Pure-AQ C18 phase (Dr. Maisch GmbH, 1.9 µm particle size). Tryptic digests were separated using a 60 min LC gradient operated at 200 nL/min. MS1 precursor scan (m/z 350–1500) acquisition was performed in the orbitrap using a nominal resolution of 30,000, followed by HCD-MS2 and ETD-MS2 fragmentation of the five

most abundant multiply charged precursor ions. Data were processed using the Sequest HT node of the Proteome Discoverer 1.4 software (Thermo Fisher Scientific). Spectra matched to glycosylated peptides were inspected manually to verify the accuracy of the assignments.

Computational docking analysis

Using the crystal structures of the EC1-EC3 regions from PcdhαC2, Pcdhβ1, PcdhγA8, and PcdhγC5 determined here, we produced docking models of Pcdh *trans* homodimers. As the results from AUC showed that the EC1-EC4 domain region self-associates to form dimers, we used the M-zdock program to generate Pcdh homodimers. For each of the crystal structures we generated 1,500 docking models resulting in a total of 6,000 models. Docked models were filtered by requiring them to include the experimentally identified specificity determinant residues at the binding interface. For docked models of PcdhαC2, PcdhgA8, and PcdhγC5 isoforms, the specificity determinant residues used were: 114, 116, 301, and 302 (PcdhγC5 numbering), whereas the specificity determinant residues 117, 118 and 121 were used to filter Pcdhβ1 isoform docking models. We did not assume an interaction between these specificity determinant residues and did not apply any distance constraints. Applying these filter conditions reduced the number of docked models from thousands to 287. The second constraint required all filtered docking models to have a buried surface area of more than 1200 $Å^2$ (600 $Å^2$ per protomer) at the binding interface, which further reduced the number of docked models to 149. To identify near-native docked homodimers we implemented the structural clustering algorithm described in (Lorenzen and Zhang, 2007). Briefly, we clustered all filtered docked models by generating an all-against-all RMSD matrix that was calculated by comparing the coordinates of protomers from different homodimer models after superposing the Cα atoms of their homophilic binding partners. Clusters were then defined by selecting a representative homodimeric model

with the most near-structural neighbors as defined by RMSD below an empirically selected threshold of 8-12 Å. Once the selected cluster representative and all its near-structural neighbors were removed from the docked-model pool, the homodimeric representative model for the next cluster was defined similarly. This procedure was repeated iteratively.

Sequence and structural alignment and Homology modelling

We modeled the structures for EC1-EC3 regions of Pcdhα7, Pcdhα8, Pcdhβ6, and Pcdhβ8 using Modeller (Eswar et al., 2006) with the crystal structures of PcdhαC2 and Pcdhβ1 as structural templates. Regions containing insertions relative to the templates were built using the LOOPY program (Soto et al., 2008). Structural alignments were calculated using DALI (Holm and Park, 2000) and SKA (Yang and Honig, 2000). Clustal Omega (Sievers et al., 2011) was used to calculate multiple sequence alignments.

Table S1: Data-collection and processing statistics for X-ray structures, related to Figure 1

	γC5 EC1-3 (Native)	γC5 EC1-3 (K_2I_4Hg)	γC5 EC1-3 (EMP)	γC5 EC1-3 (Native)	γC5 EC1-3 (Native)	γC5 EC1-3 Ext. N-terminus	αC2 EC1-3	β1 EC1-3	γA8 EC1-3
Data collection									
Date	8/15/2013	8/15/2013	8/15/2013	7/19/2013	7/19/2013	6/14/2014	9/26/2014	10/23/2014	10/23/2014
Beamline	APS 21ID-E	APS 21ID-E	APS 21ID-E	APS 21ID-E	APS 21ID-E	APS 21ID-E	BNL X4C	APS 21ID-E	APS 21ID-E
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$C2$	$P2_1$	$P2_1$	$P2_1$	$I2_12_12_1$	$I2_12_12_1$
Cell dimensions									
a, b, c (Å)	108.637, 108.637, 96.614	108.637, 108.637, 96.614	108.637, 108.637, 96.614	190.806, 104.916, 80.066	67.188, 84.563, 109.144	51.027, 108.874, 86.612	24.990, 97.130, 147.680	74.990, 106.520, 149.350	64.060, 78.120, 167.730
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 97.03, 90	90, 106.43, 90	90, 101.61, 90	90, 94.18, 90	90, 90, 90	90, 90, 90
Resolution (Å)	30.00−2.90 (3.08-2.90)	30.00-2.90 (3.00-2.90)	30.00-3.50 (3.62-3.50)	40.00-3.10 (3.21-3.10)	40.00-3.00 (3.11-3.00)	108.87-3.30 (3.48-3.30)	31.62-2.40 (2.53-2.40)	43.36-3.30 (3.56-3.30)	42.65-2.90 (3.08-2.90)
R_{merge}	0.10 (0.36)	0.15 (0.43)	0.17 (0.48)	0.11 (0.39)	0.07 (0.39)	0.12 (0.55)	0.08 (0.46)	0.15 (0.49)	0.08 (0.40)
$I/\sigma I$	19.8 (5.7)	13.5 (3.6)	16.0 (5.7)	10.0 (2.5)	16.8 (2.6)	9.6 (3.0)	9.1 (1.6)	4.6 (1.6)	10.2 (3.2)
Completeness (%)	99.9 (100.0)	98.5 (89.4)	99.8 (100.0)	94.4 (76.2)	99.1 (93.5)	99.0 (95.0)	95.8 (93.4)	97.9 (97.5)	98.9 (96.4)
Redundancy	14.2 (14.7)	11.7 (8.3)	14.0 (14.5)	3.5 (2.9)	3.7 (2.8)	3.8 (3.8)	3.0 (2.7)	2.6 (2.5)	3.1 (2.9)
Refinement									
Resolution (Å)	30.0−2.9			30.0-3.1	30.0-3.0	66.9-3.3	31.6-2.4	41.7-3.3	42.7-2.9
Number of reflections	13086			26871	23383	13880	26357	9005	9489
R_{work} / R_{free}	21.2 / 24.8			22.2 / 26.6	22.5 / 26.1	20.9 / 26.0	20.8 / 25.2	22.5 / 27.6	26.4 / 28.5
Number of residues									
Protein	311			938	642	640	634	316	313
Carbohydrate	5			14	8	9	5	12	2
Ion	6			18	12	15	12	6	7
Water	19			23	52	0	40	4	20
R.m.s. deviations									
Bond lengths (Å)	0.006			0.004	0.012	0.004	0.003	0.003	0.003
Bond angles (°)	1.229			0.780	0.871	0.753	0.730	0.878	0.654
Ramachandran									
Favored (%)	94.5			95.6	95.2	95.4	99.2	93.3	94.9
Allowed (%)	5.5			4.4	4.8	4.6	0.8	6.4	4.5
Outliers (%)	0.0			0.0	0.0	0.0	0.0	0.3	0.6
Wilson B	57.1			56.8	74.1	73.0	43.3	70.6	44.5
Average B	77.7			87.8	84.8	88.0	56.8	86.3	64.9
PDB ID	4ZPO			4ZPQ	4ZPP	4ZPN	4ZPM	4ZPL	4ZPS

Supplemental References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica Section D, Biological crystallography *66*, 213-221.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D *67*, 271-281.

delaFortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Method Enzymol *276*, 472-494.

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.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U., and Sali, A. (2006). Comparative protein structure modeling using Modeller. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al] *Chapter 5*, Unit 5 6.

Evans, P.R. (2007). An introduction to stereochemical restraints. Acta crystallographica Section D, Biological crystallography *63*, 58-61.

Halim, A., Carlsson, M.C., Madsen, C.B., Brand, S., Moller, S.R., Olsen, C.E., Vakhrushev, S.Y., Brimnes, J., Wurtzen, P.A., Ipsen, H.*, et al.* (2015). Glycoproteomic analysis of seven major allergenic proteins reveals novel post-translational modifications. Molecular & cellular proteomics : MCP *14*, 191-204.

Holm, L., and Park, J. (2000). DaliLite workbench for protein structure comparison. Bioinformatics *16*, 566-567.

Lorenzen, S., and Zhang, Y. (2007). Identification of near-native structures by clustering protein docking conformations. Proteins *68*, 187-194.

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.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Method Enzymol *276*, 307-326.

Otwinowski, Z., and Minor, W. (2001). Precision of data in diffraction experiments. Nato Sci Ser I Life *325*, 55-60.

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

Soto, C.S., Fasnacht, M., Zhu, J., Forrest, L., and Honig, B. (2008). Loop modeling: Sampling, filtering, and scoring. Proteins *70*, 834-843.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta crystallographica Section D, Biological crystallography *55*, 849-861.

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.

Yang, A.S., and Honig, B. (2000). An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance. Journal of molecular biology *301*, 665-678.