

1     **AN ADHESIVE INTERFACE FOR THE NON-CLUSTERED  $\delta$ 1 PROTOCADHERIN-1**  
2                                     **INVOLVED IN RESPIRATORY DISEASES.**

3

4                                     Debadrita Modak and Marcos Sotomayor\*

5                                     Email correspondence: [sotomayor.8@osu.edu](mailto:sotomayor.8@osu.edu)

6     Department of Chemistry and Biochemistry, The Ohio State University, 484 W 12<sup>th</sup> Avenue

7                                     Columbus, OH 43210

8                                     December 2018

1 **ABSTRACT (150 WORDS)**

2 Cadherins form a large family of calcium-dependent adhesive proteins involved in  
3 morphogenesis, cell differentiation, and neuronal connectivity. Non-clustered  $\delta 1$  protocadherins  
4 form a cadherin subgroup of proteins with seven extracellular cadherin (EC) repeats and  
5 cytoplasmic domains distinct from those of classical cadherins. The non-clustered  $\delta 1$   
6 protocadherins mediate homophilic adhesion and have been implicated in various diseases  
7 including asthma, autism, and cancer. Here we present X-ray crystal structures of Protocadherin-  
8 1 (PCDH1), a  $\delta 1$ -protocadherin member essential for New World hantavirus infection that is  
9 typically expressed in the brain, airway epithelium, skin keratinocytes, and lungs. The structures  
10 suggest a binding mode that involves antiparallel overlap of repeats EC1 to EC4. Mutagenesis  
11 combined with binding assays and biochemical experiments validated this mode of adhesion.  
12 Overall, these results reveal the molecular mechanism underlying adhesiveness of PCDH1 and  
13  $\delta 1$ -protocadherins, also shedding light on PCDH1's role in maintaining airway epithelial  
14 integrity, the loss of which causes respiratory diseases.

15

1 Cells are the basic units of life and selectively organize themselves to form tissues with the help  
2 of various adhesion proteins that engage in homophilic (same type) and heterophilic (different  
3 type) contacts. Cell-cell adhesion is of prime importance for the formation and maintenance of  
4 multicellular structures. Some of the main families of adhesion proteins include integrins,  
5 selectins, cadherins and members of the immunoglobulin superfamily<sup>1-4</sup>.

6  
7 Cadherins are calcium-dependent adhesion glycoproteins involved in a variety of biological  
8 processes such as cell differentiation and tissue morphogenesis<sup>5</sup>, cell signaling<sup>6-8</sup>,  
9 mechanotransduction<sup>9-11</sup>, and brain morphogenesis and wiring<sup>12-16</sup>. These proteins have a unique  
10 structure including tandem extracellular cadherin (EC) repeats<sup>5,17,18</sup>, followed by a  
11 transmembrane and a cytoplasmic domain. The highly similar EC repeats are around 100 amino  
12 acids long and feature highly conserved calcium-binding motifs that bind three calcium ions at  
13 the linker regions between the repeats.

14  
15 Non-clustered protocadherins are non-classical cadherins expressed in various tissues all over the  
16 body and have a wide range of functions. They are typically placed in two groups:  $\delta$  and  $\epsilon$   
17 protocadherins<sup>19</sup>. The  $\delta$ -protocadherins have seven ( $\delta 1$ ) or six ( $\delta 2$ ) EC repeats, a single  
18 transmembrane helix, and a cytoplasmic domain containing highly conserved motifs of unclear  
19 function (CM1, CM2 and CM3 in  $\delta 1$ ; CM1 and CM2 in  $\delta 2$ )<sup>19,20</sup>. These play a major role in  
20 neuronal tissue connectivity and brain wiring and are involved in various neurological disorders  
21 such as epilepsy, autism and schizophrenia<sup>20-22</sup>.

22  
23 Protocadherin-1 (PCDH1) is a non-clustered  $\delta 1$  protocadherin involved in respiratory diseases. It

1 was first identified by Sano and coworkers as PC42 in rat brain and retina tissue<sup>23</sup>. Later on, it  
2 was found to be mainly expressed in the airway epithelium, bronchial epithelium, lungs, skin  
3 keratinocytes and also in kidney, neural, and glial cells<sup>24-28</sup>. PCDH1 has seven EC repeats  
4 (labeled EC1 to EC7 from N to C-terminus), a single pass transmembrane domain and a  
5 cytoplasmic domain with the three conserved motifs (CM1-CM3) typical of the  $\delta 1$  subgroup.  
6 Immunofluorescence microscopy experiments showed that L cells transfected with PCDH1 DNA  
7 aggregated in a calcium-dependent manner and expressed the protein along the boundary of the  
8 cells at cellular contact sites<sup>23</sup>. These results suggested that PCDH1 mediated aggregation  
9 through calcium-dependent homophilic interactions.

10

11 PCDH1 has been associated with asthma, a chronic inflammatory disorder of the airways  
12 characterized by wheezing, coughing, breathlessness, chest-tightness, bronchial  
13 hyperresponsiveness (BHR) and obstruction to airflow. It is a widespread health problem  
14 affecting people of all ages, especially children, and is caused by interaction of genetic and  
15 environmental factors<sup>29,30</sup>. The epithelial layer of the airway mucosa and bronchia acts as a  
16 barrier and prevents the passage of allergens and pathogens. Dysfunction of this epithelial layer  
17 barrier is thought to play a role in asthma<sup>31</sup>. PCDH1 may bind bronchial epithelial cells and  
18 epithelial cells of the airway mucosa together by mediating homophilic adhesion and helping in  
19 forming the epithelial barrier. Loss of adhesion by PCDH1 or underexpression of the protein  
20 may lead to loss of epithelial integrity and pathogenesis<sup>31</sup>. However, how PCDH1 and other  
21 members of the  $\delta 1$ -protocadherin family can mediate cellular adhesion is unclear.

22

23 PCDH1 has also been found to be essential for New World hantaviruses infection<sup>32</sup>. These

1 viruses are transmitted from rodents to humans and cause a fatal respiratory disease called  
2 hantavirus pulmonary syndrome (HPS)<sup>33,34</sup>. There are no known vaccines or specific treatments  
3 for this disease in humans<sup>35</sup>. Results from *in vitro* and *in vivo* experiments indicate that direct  
4 recognition of PCDH1 EC1 by Gn/Gc glycoproteins on the virus envelope is a necessary step for  
5 infection<sup>32</sup>. However, the specific molecular mechanism by which PCDH1 EC1 interacts with  
6 the viral glycoproteins, and whether this interaction disrupts PCDH1 homophilic adhesion,  
7 remains to be elucidated.

8  
9 The molecular mechanisms underlying adhesion by classical cadherins are well established.  
10 These proteins use their extracellular domains to form an intercellular (*trans*) bond that is mostly  
11 homophilic (between identical cadherins)<sup>36,37</sup>. The *trans*-homophilic bond is formed by contacts  
12 between the most membrane-distal EC repeats (EC1-EC1) coming from adjacent cells<sup>38,36</sup>. The  
13 EC1-EC1 contacts are mediated by the exchange of an N-terminal  $\beta$ -strand containing one or  
14 two tryptophan residues (Trp2 or Trp2 and Trp4) that are docked into the hydrophobic pocket of  
15 the binding partner, thus forming a “strand-swapped” dimer.

16  
17 Non-classical cadherins lack the tryptophans involved in classical adhesion and use different  
18 binding mechanisms. Cadherin-23 (CDH23) and protocadherin-15 (PCDH15) form a  
19 heterophilic bond essential for inner ear mechanotransduction<sup>9,11</sup>. The heterophilic CDH23-  
20 PCDH15 bond involves an antiparallel “extended handshake” arrangement that uses both EC1  
21 and EC2 repeats<sup>39</sup>. In contrast, the clustered protocadherins ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) and the non-clustered  $\delta$ 2  
22 member PCDH19, all featuring six EC repeats, mediate *trans*-homophilic interactions using their  
23 first four extracellular repeats (EC1-4) arranged in an antiparallel bond, where EC1 interacts with

1 EC4, EC2 with EC3, EC3 with EC2 and EC4 with EC1<sup>40-45</sup>. Missense mutations in PCDH19  
2 that impair the formation of this bond are known to cause a form of epilepsy<sup>42</sup>.

3  
4 The calcium-dependent adhesion mechanisms for various other non-classical cadherins,  
5 including the non-clustered  $\delta 1$  protocadherins, are still poorly understood. Here we use bead  
6 aggregation assays to show that repeats EC1 to EC4 form the minimal adhesive unit for human  
7 (*Homo sapiens* [*hs*]) PCDH1. In addition, we present crystal structures of *hs* PCDH1 EC1-4 and  
8 show that an antiparallel overlapped dimer mediates *hs* PCDH1 adhesion. The discovery of this  
9 mode of adhesion, likely used by all other members of the  $\delta 1$  protocadherin family, along with  
10 our structures, might help in understanding the molecular mechanisms underlying PCDH1-  
11 related respiratory diseases such as asthma and HPS.

12

## 13 **RESULTS**

### 14 **Bead aggregation assays determine EC repeats required for adhesion**

15 Previous studies have shown that the full-length PCDH1 protein mediates adhesion in cell-based  
16 assays<sup>23</sup>. To determine which EC repeats are essential for PCDH1 homophilic adhesion we  
17 created a library of constructs from the human protein (hereafter referred to as PCDH1) by  
18 deleting EC repeats from the C and N-terminus. Fc-tagged protein constructs were produced in  
19 HEK293T cells and binding assays were done with Protein G magnetic beads (see Methods).  
20 Results from bead aggregation assays of PCDH1 EC1-7Fc showed that PCDH1's full-length  
21 extracellular domain mediates  $\text{Ca}^{2+}$ -dependent homophilic adhesion (Fig. 1a). Aggregation was  
22 abolished in the absence of  $\text{Ca}^{2+}$  (2 mM EDTA) (Supplementary Fig. 1a). Next, to identify the  
23 minimum unit of homophilic adhesion mediated by the extracellular domain of PCDH1, bead

1 aggregation assays were performed with PCDH1 EC1-6Fc, EC1-5Fc, EC1-4Fc, EC1-3Fc, EC1-  
2 2Fc, and EC2-7Fc (Fig. 1b-i). PCDH1 EC1-2Fc and EC1-3Fc showed no aggregation while  
3 PCDH1 EC1-4Fc, EC1-5Fc, EC1-6Fc, and EC1-7Fc showed  $Ca^{2+}$ -dependent aggregation (Fig. 1  
4 and Supplementary Fig. 1a-h). There was no aggregation when EC1 was truncated from the N-  
5 terminus (PCDH1 EC2-7Fc) (Fig. 1g). These experiments suggest that PCDH1 EC1-4 is the  
6 minimum unit of homophilic adhesion and that EC1 is required for adhesion.

7

### 8 **Bead aggregation and binding assays show that a disulfide-bond independent dimer** 9 **mediates adhesion**

10 In parallel to bead aggregation assays carried out with protein from HEK293T cells, we used  
11 PCDH1 EC1-4 and PCDH1 EC1-5 fragments refolded from *E. coli* inclusion bodies for binding  
12 assays that required large sample amounts. Refolding resulted in two monodisperse peaks  
13 corresponding to a monomer and a dimer mediated by an intermolecular cysteine-disulfide bond  
14 (data not shown). To test if free cysteines were necessary for PCDH1 dimerization, cysteine  
15 residues (C375 and C548) that were presumed to not be involved in intramolecular disulfide  
16 bonds were mutated to serine, thus abolishing potential PCDH1 intermolecular disulfide bonds.  
17 Bead aggregation assays performed with PCDH1 EC1-7Fc C375S C548S showed that PCDH1  
18 mediated bead aggregation independent of these cysteines (Fig. 2a-b). Therefore, to avoid  
19 confounding effects all binding assays of PCDH1 EC1-4 and EC1-5 obtained from bacteria were  
20 done with proteins carrying the C375S (EC1-4, EC1-5) and C548S (EC1-5) mutations.

21

22 Analytical ultracentrifugation (AUC) sedimentation velocity experiments with a range of  
23 concentrations of PCDH1 EC1-4 C375S showed both monomeric and dimeric states in solution.

1 The ratio of monomer to dimer was dependent on the concentration (Fig. 2c). Concentration  
2 dependent monomeric and dimeric states were also observed in sedimentation velocity  
3 experiments of PCDH1 EC1-5 C375S C548S (Fig. 2d). Samples were collected after AUC  
4 experiments and were run on SDS-PAGE, which indicated lack of significant degradation of the  
5 protein fragments (Supplementary Fig. 2a-b). Dimerization was observed for both PCDH1 EC1-  
6 4 C375S and PCDH1 EC1-5 C375S C548S fragments with the dimer dissociation constant ( $K_D$ )  
7 estimated to be less than 5  $\mu$ M. These results show that a dimer not formed by disulfide bonds  
8 mediates bead aggregation and exist in solution.

9

## 10 **X-ray crystal structures reveal overall architecture of PCDH1**

11 Human PCDH1 protein fragments produced by bacterial and mammalian cells were used for  
12 crystallization and structure determination attempts. Two crystal structures were determined  
13 through molecular replacement and fully refined. The first one, PCDH1 EC1-4bc at 2.85 Å, was  
14 obtained using bacterially produced protein. The second, PCDH1 EC1-4mc at 3.15 Å, was  
15 obtained using protein produced by mammalian cells (Fig. 3 and Table 1). Both structures had  
16 one molecule in the asymmetric unit with good-quality electron density maps that allowed for the  
17 unambiguous positioning of side chains for most of the residues (Supplementary Fig. 3a-b).  
18 These included V3 to N443 in PCDH1 EC1-4bc and V3 to V439 in PCDH1 EC1-4mc, with both  
19 missing 8 residues in their EC1 repeat, and with PCDH1 EC1-4bc missing 4 residues in its EC4  
20 repeat. The root mean square deviation (RMSD) for the  $C\alpha$  atoms among the chains of the two  
21 structures was  $\sim 1$  Å. Since protomers were highly similar, we will describe features as seen in  
22 the PCDH1 EC1-4bc structure, unless otherwise explicitly stated.

23



1 PCDH1 EC1-4 has four tandemly arranged EC repeats with similar architecture (seven  $\beta$  strands  
2 labeled A to G arranged in a Greek-key motif; Fig. 3a-b). Although PCDH1 EC1-4 is somewhat  
3 straight as compared to the classical cadherins, which are curved, the EC repeats have the same  
4 overall classical architecture. There are three calcium ions in each complete linker region (EC1-  
5 EC2, EC2-EC3 and EC3-EC4) and one calcium ion at the end of EC4. Calcium ions at complete  
6 linker regions are coordinated by highly conserved canonical calcium-binding motifs XEX<sup>BASE</sup>,  
7 DXD, DRE/DYE, XDX<sup>TOP</sup> and DXNDN (Supplementary Fig. 4). An additional calcium ion near  
8 the EC2-EC3 linker is unique and is at a crystal contact, but might be present due to purification  
9 and crystallization conditions with high calcium (Fig. 3b,d). The PCDH1 EC1-4mc structure  
10 does not have this additional calcium ion.

11  
12 Another notable feature observed in the PCDH1 EC1-4 structure is the presence of a 10-residue  
13 long (C66-C75) disulfide loop in EC1, between  $\beta$  strands E and F (Fig. 3c). Shorter disulfide  
14 loops (7 residues) have been previously seen in the clustered and  $\delta$ 2 protocadherin structures at  
15 this location. The presence of an extended insertion with eleven residues (I335-A345) within  $\beta$ -  
16 strand A in EC4, which we denote as “ $\delta$ -insertion”, is also unique and not seen in other clustered  
17 protocadherin structures (Fig. 3e and Supplementary Fig. 5). Similarly, a six-residue insertion in  
18 the DE loop of EC4 (G392-K397) is found only in PCDH1 sequences (Supplementary Fig. 5 and  
19 Supplementary Table 3). These unique features might be relevant for function, as further  
20 discussed below.

21  
22 There are three cysteine residues found in the EC1-4 sequence, two of which are involved in an  
23 intra-molecular disulfide bond (C66-C75) as discussed above. The third one, C375, is partially

1 buried in both structures (from bacterial and mammalian expressed proteins). This suggests that  
2 potential disulfide-mediated dimers observed in refolded PCDH1 EC1-4 might be rare or  
3 artefactual.

4  
5 The PCDH1 EC1-4mc structure had regions in which electron density could only be explained  
6 by the presence of glycosylation. Residues N89, N248 and N346 are predicted to be N-linked  
7 glycosylation sites and sugars can be observed at N248 in the PCDH1 EC1-4mc structure. C-  
8 linked glycosylation has been predicted for W186 but is not observed in our structure, while O-  
9 linked glycosylation has not been predicted but it is observed at S430 (Supplementary Fig. 3b).  
10 The role of these glycosylation sites in PCDH1 function is unclear.

11  
12 **Antiparallel interfaces in the crystal structure of PCDH1 EC1-4 suggest adhesive**  
13 **mechanisms**

14 Relevant adhesive interfaces have been observed in crystal structures of classical cadherins, of  
15 the protocadherin-15 and cadherin-23 heterophilic complex, and of clustered and non-clustered  
16 protocadherins<sup>18,39,40,42,43,46,47</sup>. The crystal structures of PCDH1 EC1-4 obtained from both  
17 bacterial and mammalian cells reveal the same two plausible interfaces of homophilic interaction  
18 (Fig. 4). The first homophilic adhesion interface (PCDH1-I1) involves two PCDH1 EC1-4  
19 molecules having complete antiparallel overlap. In this arrangement, EC1 of one molecule  
20 interacts with EC4 of the second molecule (EC1:EC4), EC2 with EC3 (EC2:EC3), EC3 with  
21 EC2 (EC3:EC2) and EC4 with EC1 (EC4:EC1) (Fig. 4a-b). The second possible interface of  
22 adhesion (PCDH1-I2) involved the opposite side of PCDH1 EC1-4, having antiparallel overlap  
23 of EC2 of one molecule with EC4 of the other (EC2:EC4), EC3 with EC3 (EC3:EC3), EC4 with

1 EC2 (EC4:EC2) and potential EC1:EC5 contacts (with EC5 not seen in the structure; Fig. 4c-d).  
2 These two possible interfaces were analyzed and tested to determine which one mediates  
3 adhesion *in vitro* and potentially *in vivo*.

4  
5 Analyses of the interaction interfaces with the Protein, Interfaces, Surfaces and Assemblies  
6 (PISA)<sup>48</sup> server revealed large interface areas of 1,539 Å<sup>2</sup> for PCDH1-I1 (385 Å<sup>2</sup> per repeat; 486  
7 Å<sup>2</sup> for EC1:EC4 and EC4:EC1, 231 Å<sup>2</sup> for EC2:EC3 and EC3:EC2; Table 2) and 1,270 Å<sup>2</sup> for  
8 PCDH1-I2 (423 Å<sup>2</sup> per repeat; 461 Å<sup>2</sup> for EC2:EC4 and EC4:EC2, 190 Å<sup>2</sup> for EC3:EC3)  
9 respectively. Both interface areas are greater than 856 Å<sup>2</sup>, an empirical cut-off to distinguish  
10 biological interfaces from crystal contacts. Therefore both are unlikely to be crystal packing  
11 artifacts<sup>49</sup>. Interestingly, the interface area per EC repeat is larger for PCDH1-I2.

12  
13 Glycosylation sites, if present at the interface, may interfere with binding and thus reveal non-  
14 physiological interfaces as in the case of VE-cadherin<sup>50</sup>. Analyses of the N- and C-linked  
15 predicted glycosylation sites showed that residues N89 and N346 are not present in either of the  
16 interfaces. In contrast, residues N248 and W186 are near the PCDH1-I2 interface. The observed  
17 O-linked glycosylated residue S430 is also near the PCDH1-I2 interface. However, sugars  
18 observed at N248 and S430 did not interfere with this interface in the crystal.

19  
20 Similarly, intermolecular disulfide bonds could interfere with binding thus favoring one interface  
21 over the other. Cysteine C375 in EC4 is located near the PCDH1-I2 interface (EC4:EC2),  
22 suggesting that an intermolecular disulfide bond at this location would block this interface and  
23 favor PCDH1-I1. Yet involvement of C375 in disulfide bond formation is not required for

1 adhesion and might be irrelevant *in vivo*.

2

3 Next, we designed mutations predicted to break each interface as determined from an analysis of  
4 the structures. A residue was considered to be at the interface if its buried surface area was at  
5 least 20% of their accessible surface area according to PISA. Residues K398 in the PCDH1-I1  
6 interface and V115 in the PCDH1-I2 interface met this criterion and were selected for  
7 mutagenesis. For the PCDH1-I1 interface, the K398 residue was found to form a salt bridge with  
8 E80 of the adjacent molecule. Therefore, we introduced the K398E mutation to prevent the  
9 interaction and the formation of the PCDH1-I1 interface. For the PCDH1-I2 interface, V115 was  
10 found to form hydrophobic interactions with V420, and therefore we introduced the V115R  
11 mutation to disrupt this interface (Fig. 5a,d).

12

13 Bead aggregation assays of PCDH1 EC1-7Fc K398E revealed no aggregation of beads in the  
14 presence of calcium (Fig. 5b,h,i). Consistent with these results, sedimentation velocity AUC  
15 experiments with protein fragments produced in bacteria and with mutated cysteine residues to  
16 avoid confounding effects showed monomeric states in solution for the K398E mutant (PCDH1  
17 EC1-4 C375S K398E and PCDH1 EC1-5 C375S C548S K398E; Fig. 5c,g and Supplementary  
18 Fig. 6a-b). In contrast, binding assays of PCDH1 EC1-7Fc V115R revealed aggregation of beads  
19 in the presence of calcium and AUC experiments showed both monomeric and dimeric states in  
20 solution for this mutant (PCDH1 EC1-5 C375S C548S V115R) (Fig. 5e-i). Melting experiments  
21 of the cysteine mutated PCDH1 EC1-5 C375S C548S, EC1-5 C375S C548S K398E and EC1-5  
22 C375S C548S V115R showed that their stability is comparable (Supplementary Fig. 7). These  
23 results indicate that the PCDH1-I1 interface mediates homodimerization in solution and is

1 responsible for bead aggregation *in vitro*.

2

### 3 **Model for homophilic interaction of PCDH1 and other $\delta 1$ protocadherins**

4 Mutagenesis coupled with binding assays and AUC experiments of the wild type and mutant  
5 protein fragments strongly suggest that PCDH1-I1 is the interface that mediates homophilic  
6 interaction. A schematic of this mode of interaction is shown in Fig. 5j. This *trans* antiparallel  
7 dimer interface suggests a mechanism of adhesion involving EC repeats 1 to 4. This mechanism  
8 of interaction is different from the tryptophan exchange mode between EC1 repeats of classical  
9 cadherins<sup>36</sup> and the “handshake” mechanism of interaction between protocadherin-15 and  
10 cadherin-23 involving EC1 and EC2<sup>39</sup>. It is similar to the “forearm handshake” mechanism of  
11 interaction of the  $\delta 2$  protocadherins and of the clustered protocadherins involving EC1 to  
12 EC4<sup>42,43,45</sup>. Although the overall mode of homophilic adhesion might be the same for  $\delta 1$ ,  $\delta 2$ , and  
13 clustered protocadherins, there are important differences. The EC1:EC4 contacts are greater in  
14 the  $\delta 1$ -protocadherin PCDH1, covering almost 64% of the total interface area (486 Å<sup>2</sup> for  
15 EC1:EC4 and 231 Å<sup>2</sup> for EC2:EC3). In contrast, both PCDH19 ( $\delta 2$  protocadherin) and clustered  
16 protocadherins EC2:EC3 interface areas are much larger than their EC1:EC4 contacts<sup>42,43</sup> (Table  
17 2).

18

19 An analysis of sequence conservation across species of interfacial residues revealed that residues  
20 at EC1:EC4 contacts are more conserved than the ones in EC2:EC3 sites (Fig. 6a-b and  
21 Supplementary Figs 8 and 9). About 43% of the interfacial area involves hydrophobic residues,  
22 ~23% are charged, and ~34% are hydrophilic residues (Supplementary Fig. 10). The EC2:EC3  
23 contacts are more amphiphilic with ~42% hydrophobic, ~17% charged, and ~42% hydrophilic

1 contacts as compared to EC1:EC4 contacts which are ~43% hydrophobic, ~26% charged and  
2 ~30% hydrophilic. All salt-bridge interactions are present in the EC1:EC4 interface. Thus, the  
3 size and charged nature of the PCDH1 EC1:EC4 interface suggest that it is important for both  
4 strength and specificity of PCDH1 homophilic adhesion.

5  
6 An interesting structural aspect of the PCDH1 EC1:EC4 interface is the involvement of the  $\delta$   
7 insertion within  $\beta$ -strand A of EC4. This eleven-residue insertion contributes four residues that  
8 directly contact EC1 of the opposite protomer (Fig. 6c). Similar insertions are found in other  
9 mouse and human  $\delta$ 1 protocadherins and the  $\delta$ 2 protocadherin-8, but not in clustered  
10 protocadherins. The  $\delta$  insertion may modulate the affinity of the interaction by providing  
11 additional contacts specific to each type of  $\delta$  protocadherin.

12  
13 Given the sequence similarity among  $\delta$ 1 protocadherins (~50%) (Supplementary Figs 11-14), the  
14 mode of interaction is likely to be the same for all other  $\delta$ 1 family members – PCDH7, PCDH9,  
15 and PCDH11X/Y. In support of this model, a very recent study showed that PCDH9 also uses its  
16 EC1-4 repeats to mediate homophilic adhesion<sup>51</sup>. A general mode of interaction in which all  $\delta$ 1  
17 protocadherins, with 7 EC repeats, use the same EC1-4 interface used by the  $\delta$ 2 and clustered  
18 protocadherins (6 EC repeats) is appealing, but rises the question of how the longer  $\delta$ 1  
19 protocadherins span the intercellular space when compared to the shorter  $\delta$ 2 and clustered  
20 protocadherins complexes. Interestingly, PCDH1 EC1-4 is more curved than the clustered  
21 protocadherins and PCDH19 EC1-4 fragments (Supplementary Fig. 15 and Videos 1 and 2).  
22 Comparison of the PCDH1-I1 complex with those formed by  $\delta$ 2 and clustered protocadherins  
23 also shows that the PCDH1 arrangement of protomers is more twisted, yet significant tilting

1 would still be required to span the same intercellular space that the other shorter members of the  
2 family can cover using the same EC1-4 binding mechanism.

3

#### 4 **DISCUSSION AND CONCLUSIONS**

5 Non-classical cadherins are part of multiple and diverse structures in multicellular organisms,  
6 including adhesive bonds important for morphogenesis, sensory perception, and neuronal  
7 connectivity<sup>11-13,36</sup>. Structural features of these cadherins are key determinants of their function.  
8 For instance, long non-classical cadherins such as protocadherin-24 (9 ECs), protocadherin-15  
9 (11 ECs), and cadherin-23 (27 ECs), are part of long protein filaments essential for enterocyte  
10 brush border morphogenesis and inner-ear mechanotransduction<sup>11,52</sup>. In contrast, the shorter  
11 clustered protocadherins, with 6 EC repeats, have been implicated in neuronal self-recognition  
12 and avoidance, and somewhat paradoxically have been shown to form weak and strictly  
13 homophilic *trans* adhesive bonds<sup>13</sup>.

14

15 In the context of brain wiring,  $\delta$  protocadherins are unique. First, unlike the clustered  
16 protocadherins, the cytoplasmic domains of  $\delta 1$  and  $\delta 2$  protocadherins have distinct sequence  
17 motifs that suggest a more diverse and versatile role in cellular signaling<sup>19</sup>. Second, a unique  $\delta$   
18 insertion in EC4 hints at a specialized role for this repeat in  $\delta$  protocadherins. Last, the length of  
19  $\delta 1$  protocadherins' extracellular domains, with 7 EC repeats, suggest a distinct set of functional  
20 roles yet to be determined. Given these differences and their differential expression in the  
21 brain<sup>20,26,27,53,54</sup>, an important question about  $\delta$  protocadherins' function pertains to whether they  
22 are involved in neuronal adhesion and circuitry specificity or rather mediate neuronal self-  
23 recognition and avoidance as proposed for the clustered protocadherins.

1  
2 Functional studies of PCDH19 indicate that this  $\delta 2$  protocadherin is required for the formation of  
3 neuronal columns in the fish optic tectum, thus implying an adhesive role<sup>55</sup>. Structural studies,  
4 validated with analyses of epilepsy-causing mutations, indicate that PCDH19 uses an extensive  
5 but weak antiparallel EC1-4 adhesive interface ideally suited to encode specificity<sup>42</sup>. Clustered  
6 protocadherins use the same antiparallel EC1-4 interface<sup>40,43-45</sup>, and sequence and evolutionary  
7 analyses suggest that at least some subgroups of clustered protocadherins use their EC1:EC4  
8 contacts to modulate the bond affinity and their EC2:EC3 interactions to determine specificity<sup>45</sup>.

9  
10 Interestingly, our structures of PCDH1 and associated binding assays indicate that this  $\delta 1$   
11 protocadherin uses an antiparallel EC1-4 interface that is similar to that used by clustered and  $\delta 2$   
12 protocadherins, yet interactions at the EC2:EC3 contacts are minimal, and salt-bridges possibly  
13 encoding for specificity are at the much larger EC1:EC4 interface. Binding assays of PCDH9,  
14 another  $\delta 1$  protocadherin, suggest that this protein also uses an antiparallel EC1-4 interface, with  
15 the EC1:EC4 contact playing a more important role than EC2:EC3 interactions in bead  
16 aggregation<sup>51</sup>. Taken together, these results indicate that  $\delta 1$  protocadherins use a unique and  
17 enlarged EC1:EC4 contact both to strengthen the adhesive bond and to determine its specificity.  
18 Whether  $\delta 1$  protocadherins use this bond to trigger a repulsive signal mediating neuronal self-  
19 recognition, to form true adhesive and specific intercellular contacts, or both depending on  
20 molecular and cellular context, is still unclear.

21  
22 The versatility of PCDH1 in mediating adhesion and signaling might be enhanced by the  
23 existence of various isoforms that modify its extracellular and cytoplasmic domains<sup>56</sup>. The



1 possible existence of a PCDH1 isoform with a shorter extracellular domain starting at EC4  
2 (NM\_001278615.1 / NP\_001265544.1) indicates that a pseudo-heterophilic bond could be  
3 formed and be mediated by the EC1:EC4 interface alone. Surface expression levels of this  
4 isoform could modulate the strength of the adhesion between two cells expressing PCDH1.

5  
6 The discovery of an antiparallel EC1-4 interface for PCDH1 and  $\delta$ 1 protocadherins with 7 ECs  
7 poses an interesting set of questions regarding the arrangement of this adhesive complex. The  
8 intercellular space occupied by classical cadherins (5 ECs interacting through EC1s in a curved  
9 and tilted arrangement,  $\sim 25$  nm)<sup>47</sup> or by clustered and  $\delta$ 2 protocadherins (6 ECs with an EC1-4  
10 overlap,  $< 40$  nm)<sup>42,44,45</sup> might be too short to accommodate a straight  $\delta$ 1 protocadherin complex  
11 that uses the EC1-4 antiparallel interface and that has 7 ECs per protomer ( $< 50$  nm). While  
12 PCDH1 is curved, and the complex is twisted, this might shorten the end-to-end distance of the  
13 bond only by a few nanometers, not enough to accommodate two additional ECs ( $\sim 10$  nm) at an  
14 intercellular space that could be dominated by shorter clustered and  $\delta$ 2 protocadherins. Tilting of  
15 the complex might solve this conundrum. It is also possible that the longer  $\delta$ 1 protocadherins are  
16 used to explore and form initial transient cell-cell contacts while other shorter cadherins and  
17 protocadherins are recruited to enable more permanent adhesion. On the contrary, a longer and  
18 straight  $\delta$ 1 complex might prevent adhesion mediated by shorter and tilted classical complexes.  
19 Determining how protocadherins orient with respect to the membrane plane, as recently explored  
20 for protocadherin-15<sup>57</sup>, will be critical to elucidate their function.

21  
22 The similarities in the overall mode of adhesion for  $\delta$ 1,  $\delta$ 2 and clustered protocadherins suggest  
23 that they may have a common ancestry<sup>58</sup>. It is thus tempting to speculate that  $\delta$  protocadherins

1 may engage in *cis* interactions as the clustered protocadherins do using their EC5 and EC6  
2 repeats<sup>41,59</sup>. Whether EC7 in  $\delta 1$  protocadherins is also involved in this type of contacts is  
3 unknown. Interestingly, PCDH19, a  $\delta 2$  protocadherin, interacts with the classical N-  
4 cadherin<sup>42,60,61</sup>. Perhaps PCDH1 interacts with E-cadherin in a similar fashion, as they co-  
5 localize in airway epithelial cells. Since the protocadherins are known to mediate weak  
6 interactions, *cis* dimerization with classical members of the family may help in forming more  
7 robust adhesion complexes.

8  
9 PCDH1 is also expressed in several non-neuronal tissues, including the cochlea, skin  
10 keratinocytes, lungs, and airway epithelia, all of which are subjected to mechanical  
11 stimuli<sup>25,28,31,62-64</sup>. The twisted adhesive bond of PCDH1 may provide some elasticity to these  
12 tissues, and its length might allow for additional stretching of cell-cell contacts without complete  
13 rupture of the adhesive contact. In addition, PCDH1 is known to help in forming the airway  
14 epithelial barrier by mediating adhesion between epithelial cells, the dysfunction of which leads  
15 to pathogenesis<sup>31</sup>. PCDH1 expression increases during differentiation of airway and bronchial  
16 epithelial cells, and it has been localized to cell-cell contact sites along with E-cadherin<sup>31</sup>. Our  
17 structural model of PCDH1 homophilic adhesion partially reveals how the airway epithelial  
18 barrier is formed, and suggests mutation sites that can be used to disrupt PCDH1 adhesion,  
19 which should help to elucidate the roles played by PCDH1's extracellular and cytoplasmic  
20 domains *in vivo*.

21  
22 The structure of PCDH1 EC1-4 and the homophilic binding mechanism presented here also have  
23 important implications in relation to New World hantaviruses infectivity. PCDH1 EC1 has been

1 recently shown to be critical for cellular attachment and entry of these viruses<sup>32</sup>, yet molecular  
2 details are still unclear. It is possible that the hantavirus glycoproteins interact with PCDH1 EC1  
3 and disrupt homophilic binding by preventing contacts with the partnering PCDH1 EC4. In this  
4 scenario, PCDH1 EC4 could be used and engineered to block interactions with the virus. On the  
5 other hand, native PCDH1 EC1 interactions with hantavirus glycoproteins may involve residues  
6 not engaged in homophilic binding, in which case PCDH1 EC1 itself could be used to develop  
7 hantavirus blockers. In either case, the molecular details provided by our PCDH1 EC1-4  
8 structure may guide drug discovery strategies aimed at reducing or blocking hantavirus  
9 infections.

10

## 11 **METHODS**

### 12 **Cloning and Mutagenesis**

13 Sequences encoding for human PCDH1 repeats EC1-4 and EC1-5 were subcloned using the  
14 NdeI and XhoI sites of the pET21a vector for bacterial expression (without signal peptide), and  
15 using the XhoI and BamHI sites of the CMV:N1-Fc and CMV:N1-His vector<sup>42</sup> for mammalian  
16 expression. Constructs encoding for truncated versions of PCDH1 (PCDH1EC1-7, PCDH1EC1-  
17 6, PCDH1EC1-5, PCDH1EC1-4, PCDH1EC1-3, PCDH1EC1-2, PCDH1EC2-7) were created by  
18 PCR subcloning of the native signal peptide and appropriate EC domains into CMV:N1-Fc.  
19 Truncation boundaries for EC repeats are as in Supplementary Fig. 4. Mutations were created in  
20 both the bacterial and the mammalian expression constructs by site-directed mutagenesis using  
21 the Quikchange Lightning Site-directed Mutagenesis Kit (Agilent). All constructs were sequence  
22 verified.

23

## 1 **Bacterial Expression, Purification, and Refolding**

2 PCDH1 fragments were expressed in BL21 CodonPlus (DE3)-RIPL cells (Stratagene) that were  
3 cultured in LB (EC3-4, EC1-4 and EC1-5 WT and mutants), induced at  $OD_{600} \sim 0.6$  with 200  $\mu$ M  
4 IPTG, and grown at 30 °C for  $\sim 18$  hr. Cells were lysed by sonication in denaturing buffer (20  
5 mM TrisHCl [pH 7.5], 6 M guanidine hydrochloride, 10 mM  $CaCl_2$  and 20 mM imidazole). The  
6 cleared lysates were loaded onto Ni-Sepharose (GE Healthcare) and eluted with denaturing  
7 buffer supplemented with 500 mM imidazole. PCDH1 fragments for crystallography were  
8 refolded by overnight dialysis against 20 mM TrisHCl [pH 8], 150 mM KCl, 50 mM NaCl, 400  
9 mM arginine, 2 mM  $CaCl_2$ , and 2 mM DTT using MWCO 2000 membranes at 4 °C. Refolded  
10 proteins were further purified by size exclusion chromatography on a Superdex200 column (GE  
11 Healthcare) in 20 mM TrisHCl [pH 8.0], 150 mM KCl, 50 mM NaCl, 2 mM  $CaCl_2$  and 2 mM  
12 DTT. PCDH1 fragments used for analytical ultracentrifugation experiments were refolded and  
13 purified using the same buffers without DTT.

14

## 15 **Mammalian Expression and Purification**

16 The PCDH1 EC1-4His fragment was expressed in Expi293<sup>TM</sup> cells (ThermoFischer Scientific).  
17 Transfection was done with 30  $\mu$ g of plasmid and ExpiFectamine<sup>TM</sup> 293. Media was collected on  
18 the fourth day after transfection by pelleting cells and collecting the supernatant to which 10 mM  
19  $CaCl_2$  was added. The mix was dialysed overnight against 20 mM TrisHCl [pH 8.0], 150 mM  
20 KCl, 50 mM NaCl, and 10 mM  $CaCl_2$  using MWCO 2000 membranes. The PCDH1 EC1-4His  
21 fragment was then purified by nickel affinity chromatography under native conditions with  
22 buffer containing 20 mM TrisHCl [pH 8.0], 300 mM NaCl, 10 mM  $CaCl_2$  and 20 mM imidazole.  
23 The protein was eluted with the same buffer containing 500 mM imidazole. It was further

1 purified by size exclusion chromatography on a Superdex200 column (GE Healthcare) in 20 mM  
2 TrisHCl [pH 8.0], 150 mM KCl, 50 mM NaCl, and 2 mM CaCl<sub>2</sub>.

3

#### 4 **Analytical Ultracentrifugation**

5 Sedimentation velocity experiments were performed in a ProteomeLab™XL-I analytical  
6 ultracentrifuge (Beckman Coulter). Samples were prepared by concentrating purified proteins  
7 from size exclusion chromatography using Sartorius Vivaspin 20 concentrators (MWCO 10  
8 kD). Protein concentrations were measured using a Nanodrop 2000c spectrophotometer (Thermo  
9 Scientific). Buffer containing 20 mM TrisHCl [pH 8.0], 150 mM KCl, 50 mM NaCl and 2 mM  
10 CaCl<sub>2</sub> was used as blank. All samples were run at 50,000 rpm at 25 °C and detection was by UV  
11 at 230 nm for samples with concentration < 8 μM and at 280 nm for others. Solvent density was  
12 assumed to be 1 g/cc and the partial specific volume of solute ( $\bar{v}$ ) was set to 0.73 cc/g. Data was  
13 analyzed using SEDFIT<sup>65,66</sup>.

14

#### 15 **Crystallization, Data Collection, and Structure Determination**

16 Crystals of PCDH1 EC1-4 fragments from bacterial and mammalian expression systems were  
17 grown by vapor diffusion at 4 °C by mixing equal volumes of protein (~ 6 mg/ml) and reservoir  
18 solution (0.1 M NaHEPES [pH 7.5], 0.2 M CaCl<sub>2</sub>, 15% PEG 400, 15% Glycerol for PCDH1  
19 EC1-4bc; 0.05 M TrisHCl [pH 8.5], 0.1 M MgCl<sub>2</sub>, 30% PEG550 MME for PCDH1 EC1-4mc).  
20 Crystals of PCDH1 EC3-4 fragments from bacterial expression systems were also grown by  
21 vapor diffusion at 4 °C by mixing equal volumes of protein (~ 8.5 mg/ml) and reservoir solution  
22 of 0.1 M MES [pH 6.5], 20% PEG 4000, and 0.6 M NaCl (cryo-protected with 25% glycerol).  
23 Crystals were cryo-cooled in liquid N<sub>2</sub>. X-ray diffraction data was collected as indicated in Table

1 1 and processed with HKL2000<sup>67</sup>. The PCDH1 EC1-4bc structure was determined by molecular  
2 replacement using an unfinished PCDH1 EC3-4 structure as an initial search model in  
3 PHASER<sup>68</sup>. Model building was done with COOT<sup>69</sup> and restrained TLS refinement was  
4 performed with REFMAC5<sup>70</sup>. Likewise, the PCDH1 EC1-4mc structure was determined through  
5 molecular replacement using the PCDH1 EC1-4bc structure as the initial search model in  
6 PHASER. Data collection and refinement statistics are provided in Table 1. Inspection of 2Fo-Fc  
7 electron density maps during initial stages of refinement of the PCDH1 EC1-4bc and PCDH1  
8 EC1-4mc model revealed unassigned densities at the linker regions. Ca<sup>2+</sup> was present at  
9 concentrations of at least 2 mM in the protein solution buffers, so we assigned Ca<sup>2+</sup> ions in the  
10 PCDH1 EC1-4bc model. However, the crystallization buffer for PCDH1 EC1-4mc had 0.1 M  
11 MgCl<sub>2</sub>. We tried to assign both Ca<sup>2+</sup> as well as Mg<sup>2+</sup> in the linker regions of the PCDH1 EC1-  
12 4mc model, but only the assignment of Ca<sup>2+</sup> ions was compatible with the final 2Fo-Fc map and  
13 was supported by an analysis of distances to coordinating atoms and B factor values of the ion  
14 and surrounding residue atoms. All molecular images were generated with VMD<sup>71</sup>.

15

## 16 **Bead Aggregation Assays**

17 Bead aggregation assays were modified from those described previously<sup>42,61,72,73</sup>. The PCDH1Fc  
18 fusion constructs were transfected into HEK293T cells using calcium-phosphate transfection<sup>74-</sup>  
19 <sup>76</sup>. Briefly, solution A (10 µg of plasmid DNA and 250 mM CaCl<sub>2</sub>) was added drop-wise to  
20 solution B (2X HBS) while mildly vortexing, and the final transfection solution was added drop-  
21 wise to one 100-mm dish of cultured HEK293T cells (two plates per assay). The next day, cells  
22 were rinsed twice with 1X PBS and serum-free media. Cells were allowed to grow in the serum-  
23 free media for 2 days before collecting the media containing the secreted Fc fusions. The media

1 was concentrated (Amicon 10 kD concentrators) and incubated with 1.5  $\mu$ l of protein G  
2 Dynabeads (Invitrogen) while rotating at 4 °C for 2 hr. The beads were washed in binding buffer  
3 (50 mM TrisHCl [pH 7.5], 100 mM NaCl, 10 mM KCl, and 0.2% BSA) and split into two tubes  
4 with either 2 mM EDTA or 2 mM CaCl<sub>2</sub>. Beads were allowed to aggregate in a glass depression  
5 slide in a humidified chamber for 60 min without motion, followed by 1 min and then 2 min of  
6 rocking (8 oscillations/min). Images were collected upon adding EDTA or CaCl<sub>2</sub>, after 60 min  
7 incubation, and after each rocking interval using a microscope (Nikon Eclipse Ti) with a 10x  
8 objective. Bead aggregates were quantified using ImageJ software as described previously<sup>61,72</sup>.  
9 Briefly, the images were thresholded, the area of the detected aggregate particles was measured  
10 in units of pixels, and the average size was calculated. Assays were repeated four times (except  
11 for PCDH1 EC1-7Fc and PCDH1 EC1-7Fc C375S C548S with  $n = 3$ ) from separate protein  
12 preps and their mean aggregate size ( $\pm$  SEM) at each time point was plotted.

13  
14 Western blots were performed using part of the media containing the Fc fusion proteins after the  
15 bead aggregation assay to confirm expression and secretion of the protein. The media was mixed  
16 with sample loading dye, boiled for 5 min and loaded onto SDS-PAGE gels (BioRad) for  
17 electrophoresis. Proteins were transferred to PVDF membrane (GE healthcare) and blocked with  
18 5% nonfat milk in TBS with 0.1% Tween-20 before incubating overnight at 4 °C with goat anti-  
19 human IgG (1:200 Jackson ImmunoResearch Laboratories). After several washes with 1X TBS,  
20 the blot was incubated with mouse anti-goat HRP-conjugated secondary (1:5000, Santa Cruz  
21 Biotechnology) for 1 hr at room temperature, washed with 1X TBS and developed with ECL  
22 Select Western Blot Detection (GE Healthcare) for chemiluminescent detection (Omega Lum G).

23

## 1 **Differential Scanning Fluorimetry**

2 The bacterially expressed PCDH1 EC1-5 C375S C548S, EC1-5 C375S C548S K398E, and EC1-  
3 5 C375S C548S V115R fragments were purified as described above and used for differential  
4 scanning fluorimetry (DSF)<sup>77-79</sup>. Experiments were repeated three times using protein at 0.2  
5 mg/ml for all constructs in buffer (20 mM TrisHCl [pH 8.0], 150 mM KCl, 50 mM NaCl, and 2  
6 mM CaCl<sub>2</sub>) mixed with SYPRO Orange dye (Invitrogen). Fluorescent measurements were  
7 performed in a CFXConnect RT-PCR instrument (BioRad) using hard-shell 96-well thin wall  
8 PCR plates (BioRad) while samples were heated from 10° to 95° in 0.2° steps. Melting  
9 temperatures were defined as those recorded when the normalized fluorescence reached 0.5.

10

## 11 **Sequence Analysis**

12 For analysis of PCDH1 residue conservation across species, 87 sequences were obtained from  
13 the NCBI protein database and processed manually to include the extracellular domain from EC  
14 repeats 1 to 7. The accession numbers for sequences of each of the 87 species is listed in  
15 Supplementary Table 1. These PCDH1 sequences were aligned using Clustal X<sup>80</sup> and the  
16 alignment file was put into ConSurf<sup>81</sup> to calculate relative conservation of each residue and to  
17 categorize the degree of conservation into nine color bins. Sequences for  $\delta$ 1 protocadherins *hs*  
18 PCDH1, *mm* PCDH1, *hs* PCDH7, *mm* PCDH7, *hs* PCDH9, *mm* PCDH9, *hs* PCDH11X and *mm*  
19 PCDH11X were obtained from the NCBI protein database, processed manually to include only  
20 the extracellular domain through the end of EC7, and aligned using the Clustal Omega server<sup>82</sup>.  
21 Identity and similarity were calculated using the Sequence Identity And Similarity SIAS server<sup>83</sup>.  
22 The accession numbers of the sequences are listed in Supplementary Table 2.

23



## 1 **Prediction of Glycosylation and Glycation Sites**

2 Potential PCDH1 glycosylation sites were predicted using the following servers: NetNGlyc 1.0  
3 (N-glycosylation, GlcNAc- $\beta$ -Asn), NetOGlyc 4.0 (O-glycosylation, GalNAc- $\alpha$ -Ser/Thr)<sup>84,85</sup>, and  
4 NetCGlyc 1.0 (C-glycosylation, Man- $\alpha$ -Trp)<sup>86</sup>.

## 6 **Orientation of EC Repeats and Calculation of Azimuthal Angles**

7 Relative orientation of EC repeats was determined using *mm* CDH23 (2WHV) as a reference.  
8 Repeat EC1 of *mm* CDH23 was aligned to the *z*-axis in VMD. Then, the N-terminal EC repeat  
9 for a pair of ECs of interest was aligned to *mm* CDH23 EC1 using COOT. The principal axis of  
10 the C-terminal EC repeat was then calculated using the Orient plugin in VMD. The projection of  
11 the principal axis on the *xy*-plane was plotted to identify the azimuthal angle  $\varphi$ . Repeat CDH23  
12 EC2 was used to define  $\varphi = 0^\circ$ .

## 14 **ACCESSION NUMBERS**

15 Coordinates for *hs* PCDH1 EC1-4bc and *hs* PCDH1 EC1-4mc have been deposited in the Protein  
16 Data Bank with entry codes 6BX7 and 6MGA, respectively.

## 18 **ACKNOWLEDGEMENTS**

19 We thank members of the Jontes and Sotomayor laboratories and Dr. Sharon R. Copper for  
20 training, assistance, and discussions. This work was supported in part by the Ohio State  
21 University. Use of APS NE-CAT beamlines was supported by NIH (P30 GM124165 & S10  
22 RR029205) and the Department of Energy (DE-AC02-06CH11357) through grants GUP 49774  
23 and 59251. DM is a Pelotonia fellow and MS was an Alfred P. Sloan fellow (FR-2015-65794).

1

## 2 AUTHOR CONTRIBUTIONS

3 D.M. did cloning, protein expression and protein purification, and carried out all binding and  
4 bead aggregation assays, crystallization trials, differential scanning fluorimetry experiments, and  
5 sequence alignments. M.S. trained D.M. and assisted with crystal fishing and cryo-cooling. D.M.  
6 and M.S. solved crystal structures. M.S. supervised work and assisted with data analysis. D.M.  
7 and M.S. prepared figures and wrote the manuscript.

8

## 9 DECLARATION OF INTERESTS

10 The authors declare no competing interests.

11

## 12 REFERENCES

13

- 14 1. Steinberg, M. S. Adhesion in development: an historical overview. *Dev. Biol.* **180**, 377–88  
15 (1996).
- 16 2. Shapiro, L., Love, J. & Colman, D. R. Adhesion molecules in the nervous system:  
17 structural insights into function and diversity. *Annu. Rev. Neurosci.* **30**, 451–74 (2007).
- 18 3. Hynes, R. O. Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673–87  
19 (2002).
- 20 4. Tucker, R. P. & Chiquet-Ehrismann, R. Teneurins: A conserved family of transmembrane  
21 proteins involved in intercellular signaling during development. *Dev. Biol.* **290**, 237–245  
22 (2006).
- 23 5. Takeichi, M. Morphogenetic roles of classic cadherins. *Curr. Opin. Cell Biol.* **7**, 619–627  
24 (1995).
- 25 6. Wheelock, M. J. & Johnson, K. R. Cadherin-mediated cellular signaling. *Curr. Opin. Cell*  
26 *Biol.* **15**, 509–514 (2003).
- 27 7. Suzuki, S. T. Structural and functional diversity of cadherin superfamily: Are new  
28 members of cadherin superfamily involved in signal transduction pathway? *J. Cell.*  
29 *Biochem.* **61**, 531–542 (1996).
- 30 8. Sotomayor, M., Gaudet, R. & Corey, D. P. Sorting out a promiscuous superfamily:  
31 towards cadherin connectomics. *Trends Cell Biol.* **24**, 524–36 (2014).
- 32 9. Gillespie, P. G. & Müller, U. Mechanotransduction by hair cells: models, molecules, and  
33 mechanisms. *Cell* **139**, 33–44 (2009).
- 34 10. Pruitt, B. L., Dunn, A. R., Weis, W. I. & Nelson, W. J. Mechano-transduction: from  
35 molecules to tissues. *PLoS Biol.* **12**, e1001996 (2014).

- 1 11. Jaiganesh, A., Narui, Y., Araya-Secchi, R. & Sotomayor, M. Beyond Cell-Cell Adhesion:  
2 Sensational Cadherins for Hearing and Balance. *Cold Spring Harb. Perspect. Biol.*  
3 a029280 (2017). doi:10.1101/cshperspect.a029280
- 4 12. Hirano, S. & Takeichi, M. Cadherins in Brain Morphogenesis and Wiring. *Physiol. Rev.*  
5 **92**, 597–634 (2012).
- 6 13. Mountoufaris, G., Canzio, D., Nwakeze, C. L., Chen, W. V. & Maniatis, T. Writing,  
7 Reading, and Translating the Clustered Protocadherin Cell Surface Recognition Code for  
8 Neural Circuit Assembly. *Annu. Rev. Cell Dev. Biol.* **34**, 471–493 (2018).
- 9 14. Lefebvre, J. L., Kostadinov, D., Chen, W. V, Maniatis, T. & Sanes, J. R. Protocadherins  
10 mediate dendritic self-avoidance in the mammalian nervous system. *Nature* **488**, 517–21  
11 (2012).
- 12 15. Schreiner, D. & Weiner, J. A. Combinatorial homophilic interaction between gamma-  
13 protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc.*  
14 *Natl. Acad. Sci. U. S. A.* **107**, 14893–8 (2010).
- 15 16. Weiner, J. A. & Jontes, J. D. Protocadherins, not prototypical: a complex tale of their  
16 interactions, expression, and functions. *Front. Mol. Neurosci.* **6**, 4 (2013).
- 17 17. Shapiro, L. *et al.* Structural basis of cell-cell adhesion by cadherins. *Nature* **374**, 327–37  
18 (1995).
- 19 18. Nagar, B., Overduin, M., Ikura, M. & Rini, J. M. Structural basis of calcium-induced E-  
20 cadherin rigidification and dimerization. *Nature* **380**, 360–4 (1996).
- 21 19. Redies, C., Vanhalst, K. & Roy, F. van. delta-Protocadherins: unique structures and  
22 functions. *Cell. Mol. Life Sci.* **62**, 2840–52 (2005).
- 23 20. Kim, S.-Y., Yasuda, S., Tanaka, H., Yamagata, K. & Kim, H. Non-clustered  
24 protocadherin. *Cell Adh. Migr.* **5**, 97–105 (2011).
- 25 21. Redies, C., Hertel, N. & Hübner, C. A. Cadherins and neuropsychiatric disorders. *Brain*  
26 *Res.* **1470**, 130–44 (2012).
- 27 22. Hirabayashi, T. & Yagi, T. Protocadherins in neurological diseases. *Adv. Neurobiol.* **8**,  
28 293–314 (2014).
- 29 23. Sano, K. *et al.* Protocadherins: a large family of cadherin-related molecules in central  
30 nervous system. *EMBO J.* **12**, 2249–56 (1993).
- 31 24. Bononi, J., Cole, A., Tewson, P., Schumacher, A. & Bradley, R. Chicken protocadherin-1  
32 functions to localize neural crest cells to the dorsal root ganglia during PNS formation.  
33 *Mech. Dev.* **125**, 1033–1047 (2008).
- 34 25. Koning, H. *et al.* Characterization of protocadherin-1 expression in primary bronchial  
35 epithelial cells: association with epithelial cell differentiation. *FASEB J.* **26**, 439–48  
36 (2012).
- 37 26. Blevins, C. J., Emond, M. R., Biswas, S. & Jontes, J. D. Differential expression,  
38 alternative splicing, and adhesive properties of the zebrafish  $\delta 1$ -protocadherins.  
39 *Neuroscience* **199**, 523–34 (2011).
- 40 27. Redies, C., Heyder, J., Kohoutek, T., Staes, K. & Van Roy, F. Expression of  
41 protocadherin-1 (*Pcdh1*) during mouse development. *Dev. Dyn.* **237**, 2496–2505 (2008).
- 42 28. Mortensen, L. J., Kreiner-Møller, E., Hakonarson, H., Bønnelykke, K. & Bisgaard, H. The  
43 *PCDH1* gene and asthma in early childhood. *Eur. Respir. J.* **43**, 792–800 (2014).
- 44 29. Blumenthal, J. B. & Blumenthal, M. N. Genetics of asthma. *Medical Clinics of North*  
45 *America* **86**, 937–950 (2002).
- 46 30. Bateman, E. D. *et al.* Global strategy for asthma management and prevention: GINA

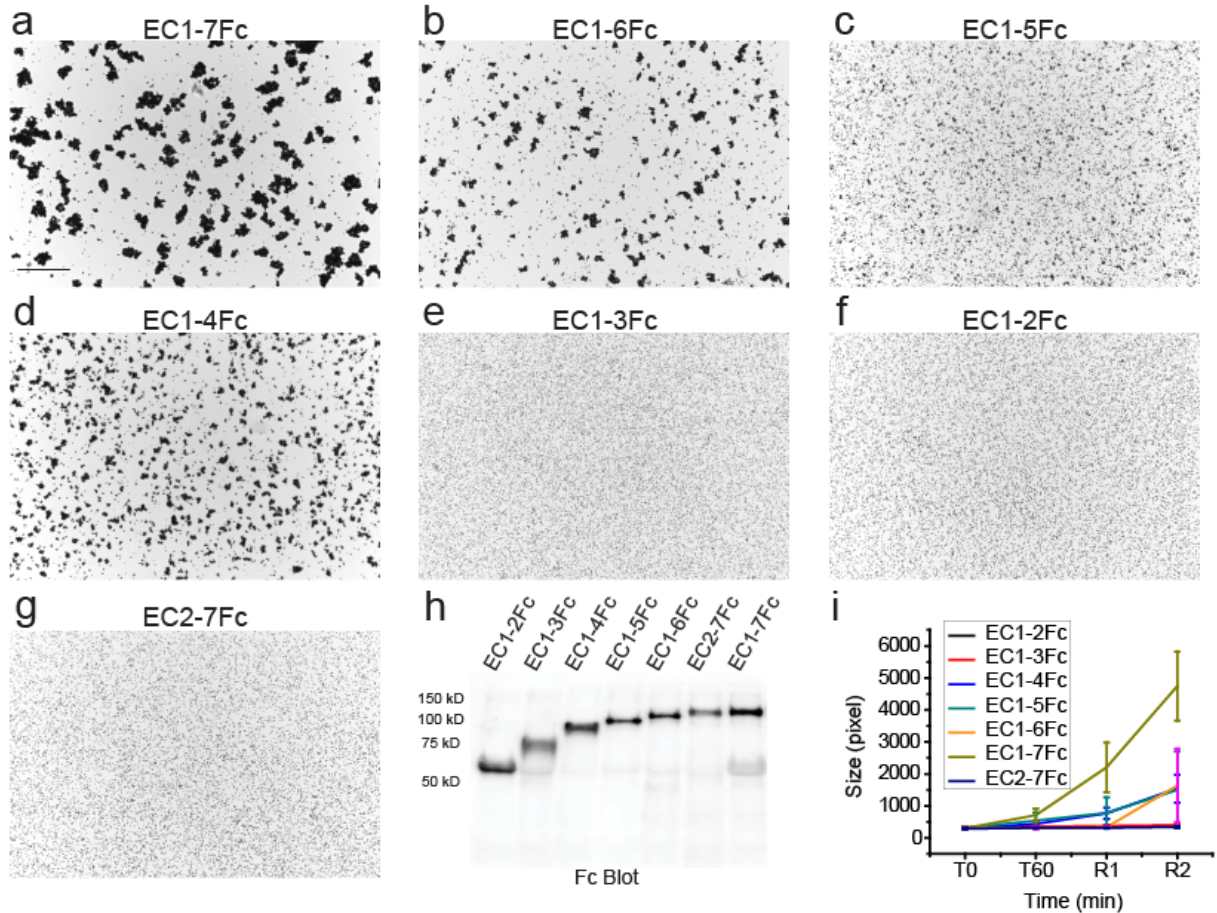
- 1 executive summary. *European Respiratory Journal* **31**, 143–178 (2008).
- 2 31. Kozu, Y. *et al.* Protocadherin-1 is a glucocorticoid- responsive critical regulator of airway  
3 epithelial barrier function. *BMC Pulm. Med.* **1**, 1–12 (2015).
- 4 32. Jangra, R. K. *et al.* Protocadherin-1 is essential for cell entry by New World hantaviruses.  
5 *Nature* **563**, 559–563 (2018).
- 6 33. Zaki, S. R. *et al.* Hantavirus pulmonary syndrome. Pathogenesis of an emerging infectious  
7 disease. *Am. J. Pathol.* **146**, 552–79 (1995).
- 8 34. Macneil, A., Nichol, S. T. & Spiropoulou, C. F. Hantavirus pulmonary syndrome. *Virus*  
9 *Res.* **162**, 138–47 (2011).
- 10 35. Garrido, J. L. *et al.* Two recombinant human monoclonal antibodies that protect against  
11 lethal Andes hantavirus infection in vivo. *Sci. Transl. Med.* **10**, eaat6420 (2018).
- 12 36. Brasch, J., Harrison, O. J., Honig, B. & Shapiro, L. Thinking outside the cell: how  
13 cadherins drive adhesion. *Trends Cell Biol.* **22**, 299–310 (2012).
- 14 37. Brasch, J. *et al.* Homophilic and Heterophilic Interactions of Type II Cadherins Identify  
15 Specificity Groups Underlying Cell-Adhesive Behavior. *Cell Rep.* **23**, 1840–1852 (2018).
- 16 38. Vendome, J. *et al.* Molecular design principles underlying  $\beta$ -strand swapping in the  
17 adhesive dimerization of cadherins. *Nat. Struct. Mol. Biol.* **18**, 693–700 (2011).
- 18 39. Sotomayor, M., Weihofen, W. A., Gaudet, R. & Corey, D. P. Structure of a force-  
19 conveying cadherin bond essential for inner-ear mechanotransduction. *Nature* **492**, 128–  
20 32 (2012).
- 21 40. Nicoludis, J. M. *et al.* Structure and Sequence Analyses of Clustered Protocadherins  
22 Reveal Antiparallel Interactions that Mediate Homophilic Specificity. *Structure* **23**, 2087–  
23 98 (2015).
- 24 41. Rubinstein, R. *et al.* Molecular Logic of Neuronal Self-Recognition through Protocadherin  
25 Domain Interactions. *Cell* **163**, 629–42 (2015).
- 26 42. Cooper, S. R., Jontes, J. D. & Sotomayor, M. Structural determinants of adhesion by  
27 Protocadherin-19 and implications for its role in epilepsy. *Elife* **5**, (2016).
- 28 43. Goodman, K. M. *et al.* Structural Basis of Diverse Homophilic Recognition by Clustered  
29  $\alpha$ - and  $\beta$ -Protocadherins. *Neuron* (2016). doi:10.1016/j.neuron.2016.04.004
- 30 44. Goodman, K. M. *et al.*  $\gamma$ -Protocadherin structural diversity and functional implications.  
31 *Elife* **5**, (2016).
- 32 45. Nicoludis, J. M. *et al.* Antiparallel protocadherin homodimers use distinct affinity- and  
33 specificity-mediating regions in cadherin repeats 1-4. *Elife* **5**, (2016).
- 34 46. Ciatto, C. *et al.* T-cadherin structures reveal a novel adhesive binding mechanism. *Nat.*  
35 *Struct. Mol. Biol.* **17**, 339–47 (2010).
- 36 47. Harrison, O. J. *et al.* The extracellular architecture of adherens junctions revealed by  
37 crystal structures of type I cadherins. *Structure* **19**, 244–56 (2011).
- 38 48. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state.  
39 *J. Mol. Biol.* **372**, 774–97 (2007).
- 40 49. Ponstingl, H., Henrick, K. & Thornton, J. M. Discriminating between homodimeric and  
41 monomeric proteins in the crystalline state. *Proteins* **41**, 47–57 (2000).
- 42 50. Brasch, J. *et al.* Structure and binding mechanism of vascular endothelial cadherin: a  
43 divergent classical cadherin. *J. Mol. Biol.* **408**, 57–73 (2011).
- 44 51. Peng, X. *et al.* Affinity capture of polyribosomes followed by RNAseq (ACAPseq), a  
45 discovery platform for protein-protein interactions. *Elife* **7**, (2018).
- 46 52. Crawley, S. W., Mooseker, M. S. & Tyska, M. J. Shaping the intestinal brush border. *J.*

- 1           *Cell Biol.* **207**, 441–51 (2014).
- 2   53.   Kim, S.-Y., Chung, H. S., Sun, W. & Kim, H. Spatiotemporal expression pattern of non-  
3       clustered protocadherin family members in the developing rat brain. *Neuroscience* **147**,  
4       996–1021 (2007).
- 5   54.   Vanhalst, K., Kools, P., Staes, K., van Roy, F. & Redies, C. delta-Protocadherins: a gene  
6       family expressed differentially in the mouse brain. *Cell. Mol. Life Sci.* **62**, 1247–59  
7       (2005).
- 8   55.   Cooper, S. R. *et al.* Protocadherins control the modular assembly of neuronal columns in  
9       the zebrafish optic tectum. *J. Cell Biol.* **211**, 807–14 (2015).
- 10   56.   Koning, H. *et al.* Mouse Protocadherin-1 Gene Expression Is Regulated by Cigarette  
11       Smoke Exposure In Vivo. *PLoS One* **9**, e98197 (2014).
- 12   57.   Ge, J. *et al.* Structure of mouse protocadherin 15 of the stereocilia tip link in complex with  
13       LHFPL5. *Elife* **7**, (2018).
- 14   58.   Oda, H. & Takeichi, M. Evolution: structural and functional diversity of cadherin at the  
15       adherens junction. *J. Cell Biol.* **193**, 1137–46 (2011).
- 16   59.   Goodman, K. M. *et al.* Protocadherin *cis* -dimer architecture and recognition unit  
17       diversity. *Proc. Natl. Acad. Sci.* **114**, E9829–E9837 (2017).
- 18   60.   Biswas, S., Emond, M. R. & Jontes, J. D. Protocadherin-19 and N-cadherin interact to  
19       control cell movements during anterior neurulation. *J. Cell Biol.* **191**, 1029–41 (2010).
- 20   61.   Emond, M. R., Biswas, S., Blevins, C. J. & Jontes, J. D. A complex of Protocadherin-19  
21       and N-cadherin mediates a novel mechanism of cell adhesion. *J. Cell Biol.* **195**, 1115–21  
22       (2011).
- 23   62.   Koppelman, G. H. *et al.* Identification of PCDH1 as a novel susceptibility gene for  
24       bronchial hyperresponsiveness. *Am. J. Respir. Crit. Care Med.* **180**, 929–935 (2009).
- 25   63.   Faura Tellez, G. *et al.* Protocadherin-1 Localization and Cell-Adhesion Function in  
26       Airway Epithelial Cells in Asthma. *PLoS One* **11**, e0163967 (2016).
- 27   64.   Lin, J. *et al.* Anatomical expression patterns of delta-protocadherins in developing chicken  
28       cochlea. *J. Anat.* **221**, 598–608 (2012).
- 29   65.   Schuck, P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity  
30       Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* **78**, 1606–1619 (2000).
- 31   66.   Schuck, P. On the analysis of protein self-association by sedimentation velocity analytical  
32       ultracentrifugation. *Anal. Biochem.* **320**, 104–24 (2003).
- 33   67.   Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the  
34       integration of data reduction and structure solution--from diffraction images to an initial  
35       model in minutes. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 859–66 (2006).
- 36   68.   McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674  
37       (2007).
- 38   69.   Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot.  
39       *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
- 40   70.   Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal  
41       structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
- 42   71.   Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol.*  
43       *Graph.* **14**, 33–38 (1996).
- 44   72.   Emond, M. R. & Jontes, J. D. Bead aggregation assays for the characterization of putative  
45       cell adhesion molecules. *J. Vis. Exp.* e51762 (2014). doi:10.3791/51762
- 46   73.   Sivasankar, S., Zhang, Y., Nelson, W. J. & Chu, S. Characterizing the initial encounter

- 1 complex in cadherin adhesion. *Structure* **17**, 1075–81 (2009).
- 2 74. Kwon, M. & Firestein, B. L. DNA transfection: calcium phosphate method. *Methods Mol.*  
3 *Biol.* **1018**, 107–10 (2013).
- 4 75. Barry, J. *et al.* Ankyrin-G directly binds to kinesin-1 to transport voltage-gated Na<sup>+</sup>  
5 channels into axons. *Dev. Cell* **28**, 117–31 (2014).
- 6 76. Jiang, M. & Chen, G. High Ca<sup>2+</sup>-phosphate transfection efficiency in low-density  
7 neuronal cultures. *Nat. Protoc.* **1**, 695–700 (2006).
- 8 77. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to  
9 detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–21 (2007).
- 10 78. Lavinder, J. J., Hari, S. B., Sullivan, B. J. & Magliery, T. J. High-throughput thermal  
11 scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *J. Am.*  
12 *Chem. Soc.* **131**, 3794–5 (2009).
- 13 79. Choudhary, D., Kumar, A., Magliery, T. J. & Sotomayor, M. Using thermal scanning  
14 assays to test protein-protein interactions of inner-ear cadherins. *PLoS One* **12**, e0189546  
15 (2017).
- 16 80. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948  
17 (2007).
- 18 81. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize  
19 evolutionary conservation in macromolecules. *Nucleic Acids Res.* (2016).  
20 doi:10.1093/nar/gkw408
- 21 82. Sievers, F. & Higgins, D. G. Clustal Omega, accurate alignment of very large numbers of  
22 sequences. *Methods Mol. Biol.* **1079**, 105–16 (2014).
- 23 83. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence  
24 alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
- 25 84. Hansen, J. E. *et al.* NetOglyc: prediction of mucin type O-glycosylation sites based on  
26 sequence context and surface accessibility. *Glycoconj. J.* **15**, 115–30 (1998).
- 27 85. Steentoft, C. *et al.* Precision mapping of the human O-GalNAc glycoproteome through  
28 SimpleCell technology. *EMBO J.* **32**, 1478–88 (2013).
- 29 86. Julenius, K. NetCGlyc 1.0: prediction of mammalian C-mannosylation sites. *Glycobiology*  
30 **17**, 868–76 (2007).
- 31
- 32

1 **FIGURES AND FIGURE LEGENDS**

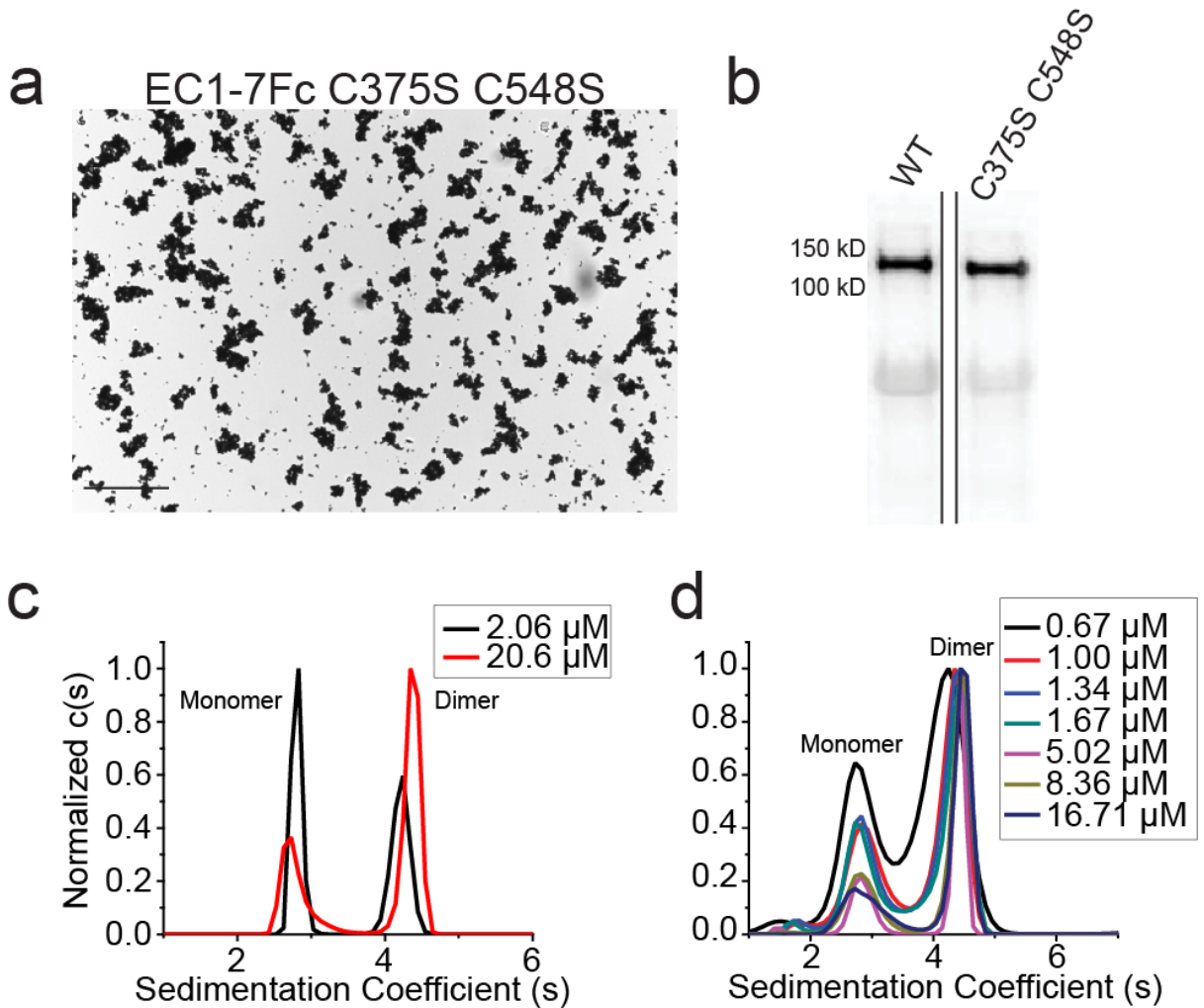
2



3

4 **Figure 1. Minimal adhesive unit of PCDH1 is EC1-4.** (a-g) Protein G beads coated with full length (a) and  
5 truncated versions (b-g) of the PCDH1 extracellular domain imaged after incubation for 1 hr followed by rocking  
6 for 2 min in the presence of calcium. Bar – 500  $\mu$ m. (h) Western blot shows efficient expression of full length and  
7 truncated PCDH1 EC repeats. (i) Mean aggregate size for full length and truncated fragments of PCDH1 at T0 ( $t = 0$   
8 min), after 1 hr of incubation, T60 ( $t = 60$  min) followed by rocking for 1 min (R1) and 2 min (R2). Error bars are  
9 standard error of the mean ( $n = 4$  for all aggregation assays and constructs except for PCDH1 EC1-7Fc;  $n = 3$ ).

1

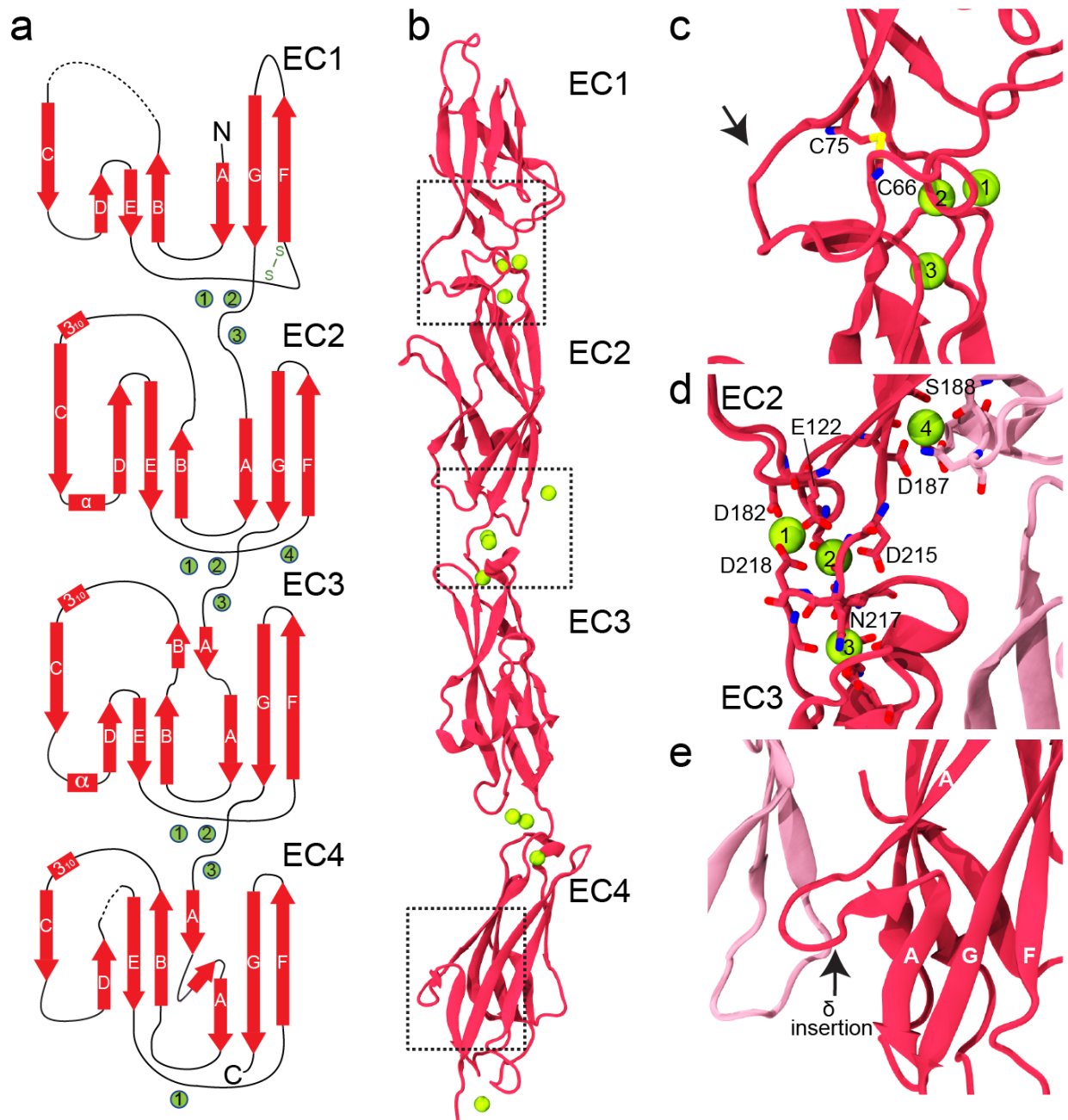


2

3 **Figure 2. PCDH1-mediated bead aggregation is independent of disulfide bond formation and PCDH1 EC1-4**  
4 **C375S and PCDH1 EC1-5 C375S C548S form dimers in solution. (a)** Protein G beads coated with the PCDH1  
5 EC1-7Fc C375S C548S mutant imaged after incubation for 1 hr followed by rocking for 2 min in the presence of  
6 calcium. Bar – 500  $\mu$ m. Bead aggregation is evident. **(b)** Western blot shows efficient expression of wild type  
7 PCDH1 EC1-7Fc and mutant PCDH1 EC1-7Fc C375S C548S. **(c)** AUC data from bacterially produced PCDH1  
8 EC1-4 C375S show both dimeric and monomeric peaks at 2.06  $\mu$ M and 20.6  $\mu$ M with normalized c(s) plotted  
9 against sedimentation coefficient (s). **(d)** AUC data from bacterially produced PCDH1 EC1-5 C375S C548S show  
10 both dimeric and monomeric peaks for a range of concentrations between 0.67  $\mu$ M and 16.71  $\mu$ M.



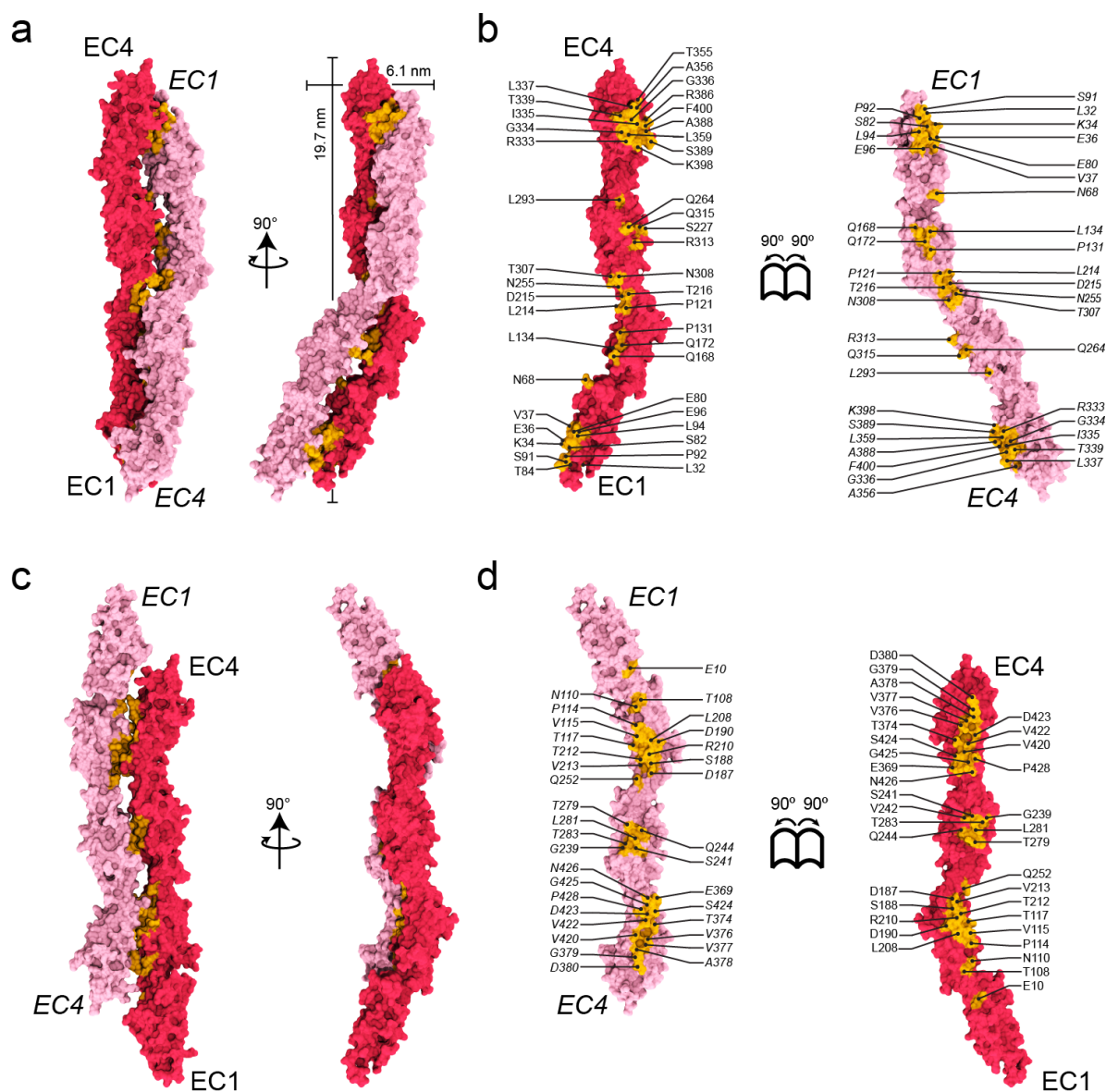
1



2

3 **Figure 3. X-ray crystal structure of PCDH1 EC1-4.** (a) Topology diagram of PCDH1 EC1-4bc. A typical  
4 cadherin fold is observed for each EC repeat with seven  $\beta$  strands labeled A to G. Calcium ions are shown as green  
5 circles. Dashed lines are missing loops. (b) Ribbon diagram of PCDH1 EC1-4bc (6BX7). Black dashed boxes  
6 indicate unique features of PCDH1 EC1-4 highlighted in panels (c-e). (c) Details of a disulfide loop in EC1 (arrow).  
7 (d) Details of the calcium-binding site in the EC2-EC3 linker region. (e) Details of an extended  $\delta$  insertion within  $\beta$   
8 strand A in EC4 not seen in clustered protocadherins (arrow).

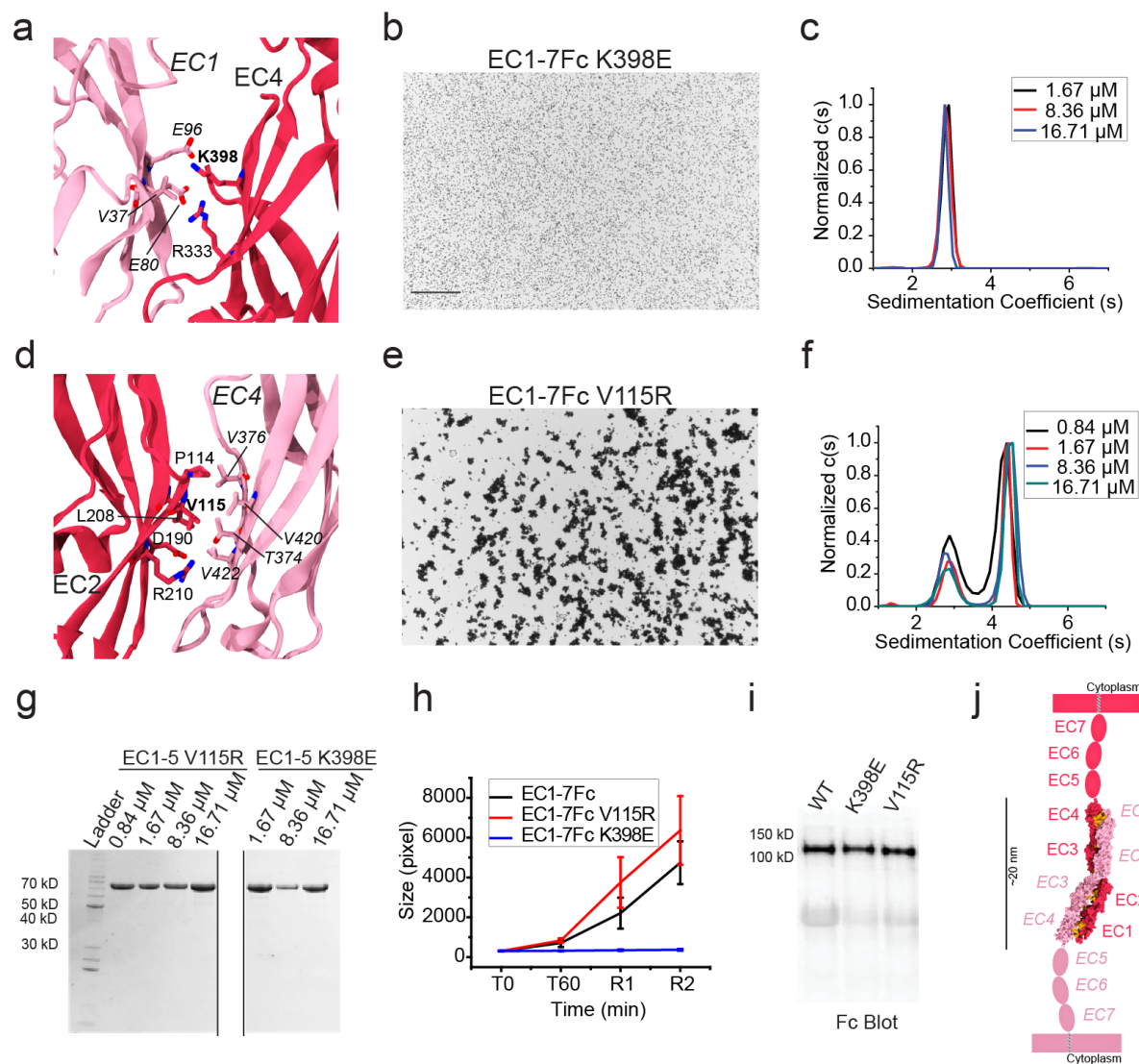
1



2  
3 **Figure 4. Crystallographic antiparallel interfaces of PCDH1.** (a) Molecular surface representation of two  
4 PCDH1 EC1-4 molecules in the crystal with the interaction interface formed by fully overlapped and antiparallel  
5 EC1-4 repeats (PCDH1-I1). Two perpendicular views are shown. (b) Interaction surface exposed with interfacing  
6 residues listed and shown in yellow. (c) Molecular surface representation of two PCDH1 EC1-4 molecules in an  
7 alternate crystallographic arrangement in which the interaction interface would be formed by fully overlapped and  
8 antiparallel EC1-5 repeats (PCDH1-I2). Two perpendicular views are shown. (d) Interaction surface exposed with  
9 interfacing residues listed and shown in yellow.

10

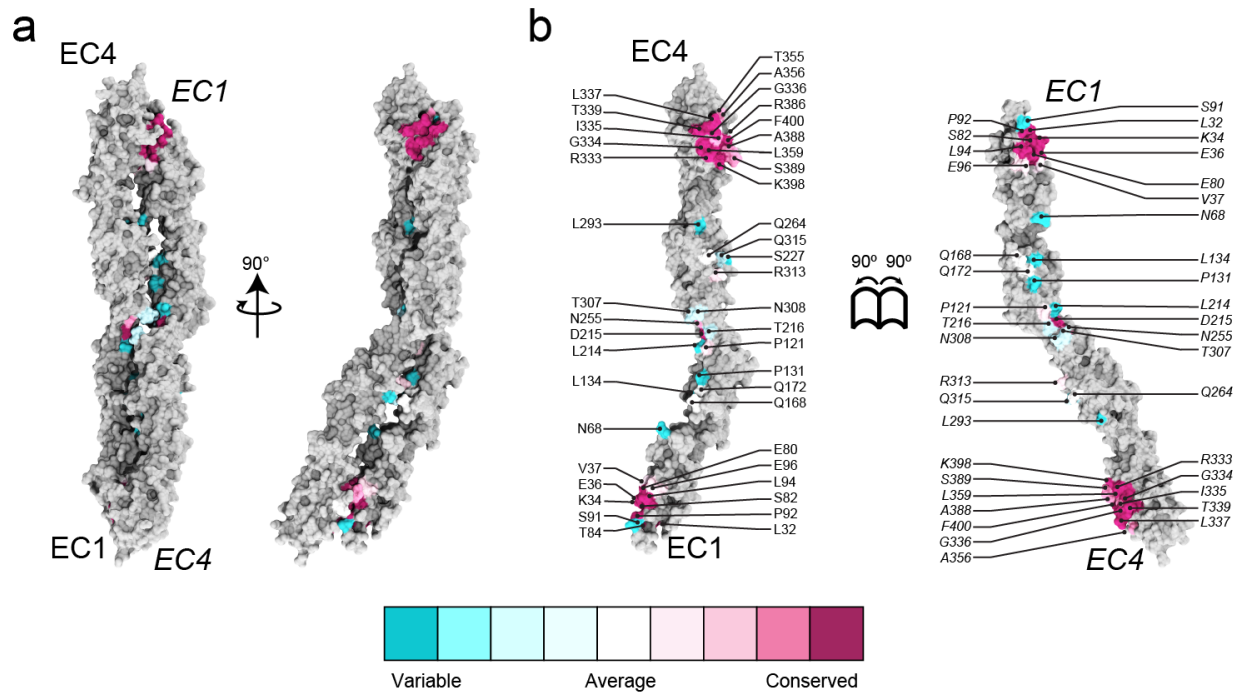
1



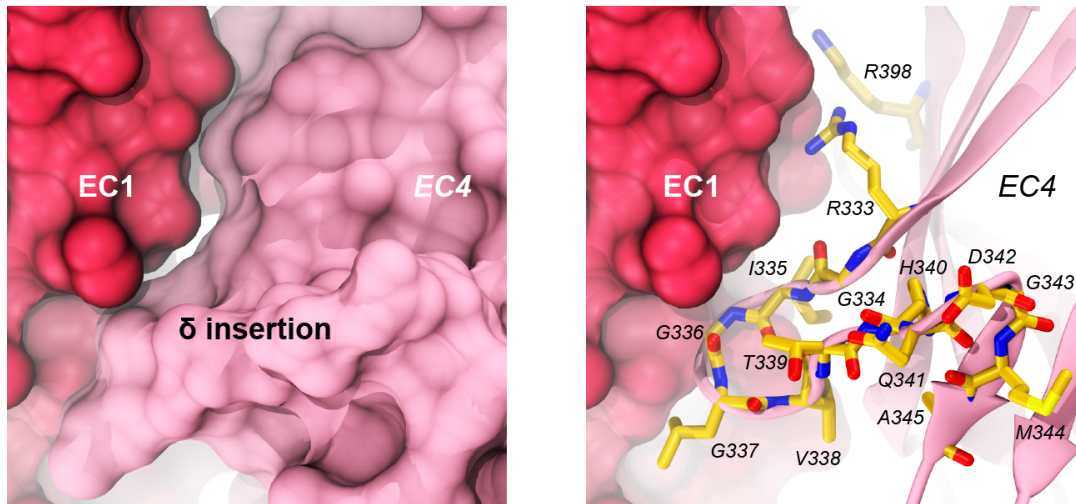
2

3 **Figure 5. Testing of crystallographic PCDH1 antiparallel interfaces.** (a) Detail of the PCDH1-I1 interface  
 4 (EC1:EC4). Residue K398E interacting with E80 is highlighted. (b) Protein G beads coated with PCDH1 EC1-7Fc  
 5 K398E mutant protein and imaged after incubation for 1 hr followed by rocking for 2 min in the presence of  
 6 calcium. Bar – 500  $\mu$ m. No aggregation is observed. (c) AUC of bacterially produced PCDH1 EC1-5 C375S C548S  
 7 K398E shows only a monomeric peak for a range of concentrations between 1.67  $\mu$ M and 16.71  $\mu$ M. (d) Detail of  
 8 the PCDH1-I2 interface (EC1:EC5). Residue V115 interacting with V420 is highlighted. (e) Protein G beads coated  
 9 with PCDH1 EC1-7Fc V115R mutant protein and imaged after incubation for 1 hr followed by rocking for 2 min in  
 10 the presence of calcium. Bar – 500  $\mu$ m. Aggregation is observed. (f) AUC of bacterially produced PCDH1 EC1-5  
 11 C375S C548S V115R shows both dimeric and monomeric peaks for a range of concentrations between 0.84  $\mu$ M and  
 12 16.71  $\mu$ M. (g) SDS-PAGE shows samples for PCDH1 EC1-5 C375S C548S V115R and PCDH1 EC1-5 C375S  
 13 C548S K398E after AUC. Different amounts of samples were loaded while running the gel. (h) Mean aggregate size  
 14 for wild type PCDH1 EC1-7Fc and mutants PCDH1 EC1-7Fc K398E, PCDH1 EC1-7Fc V115R at T0 ( $t = 0$  min),  
 15 after 1 hr of incubation, T60 ( $t = 60$  min) followed by rocking for 1 min (R1) and 2 min (R2). Error bars are standard  
 16 error of the mean ( $n = 4$  for mutants,  $n = 3$  for wild type). (i) Western blot shows efficient expression of wild type  
 17 PCDH1 EC1-7Fc and mutants PCDH1 EC1-7Fc K398E and PCDH1 EC1-7Fc V115R. (j) Schematic of *trans*  
 18 antiparallel binding of PCDH1 involving overlap of repeats EC1 to EC4.  
 19

1



**c**



2

3 **Figure 6. Conservation of residues at the crystallographic PCDH1-I1 interface.** (a) Molecular surface  
 4 representation of PCDH1-I1 antiparallel dimer. Two perpendicular views are shown. (b) Interaction surface exposed  
 5 with interfacial residues listed and colored according to sequence conservation among 87 species (right). The  
 6 accession numbers for the sequences of 87 species are given in Supplementary Table 1. (c) Detail of the  $\delta$  insertion  
 7 in EC4 shown in molecular surface representation (left) and in ribbon with relevant residues as sticks (right).  
 8

1 **TABLES**

2

3 **Table 1**

4

5 Statistics for *hs* PCDH1 structures.

<b>Data collection</b>	<b><i>Hs</i> PCDH1 EC1-4bc</b>	<b><i>Hs</i> PCDH1 EC1-4mc</b>	<b><i>Hs</i> PCDH1 EC3-4</b>
Space group	P6 <sub>1</sub> 22	P6 <sub>1</sub> 22	P4 <sub>2</sub> 2 <sub>1</sub> 2
Unit cell parameters			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	146.98, 146.98, 147.69	147.23, 147.23, 149.37	125.57, 125.57, 45.29
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 120	90, 90, 120	90, 90, 90
Molecules per asymmetric unit	1	1	1
Beam source	24-ID-E	24-ID-E	24-ID-E
Date of data collection	18-DEC-2016	15-JUN-2017	30-JUL-2016
Wavelength (Å)	0.97921	0.97918	0.97921
Resolution (Å)	2.85	3.15	3.05
Unique reflections	22567	17231	7363
Completeness (%)	100.0 (100.0)	100.0 (99.9)	99.3 (90.6)
Redundancy	10.7 (10.6)	36.4 (28.2)	9.9 (6.5)
<i>I</i> / $\sigma$ ( <i>I</i> )	18.3 (1.923)	58 (2.5)	12 (1.67)
<i>R</i> <sub>merge</sub>	0.156 (1.627)	0.078 (1.762)	0.166 (1.027)
<i>R</i> <sub>meas</sub>	0.164 (1.710)	0.079 (1.792)	0.175 (1.110)
<i>R</i> <sub>pim</sub>	0.050 (0.521)	0.013 (0.321)	0.054 (0.401)
<i>CC</i> <sub>1/2</sub>	0.998 (0.716)	1.004 (0.794)	0.938 (0.551)
<i>CC</i> *	0.999 (0.913)	1.001 (0.941)	0.981 (0.843)
<b>Refinement</b>			
Resolution range (Å)	48.26-2.85 (2.93-2.85)	48.53-3.15 (3.22-3.15)	
<i>R</i> <sub>work</sub> (%)	22.1 (33.8)	21.5 (33.3)	
<i>R</i> <sub>free</sub> (%)	24.3 (37.2)	26.3 (43.0)	
Protein Residues	429	429	
Ligands/ions	12	9	
Water molecules	25	0	
Rms deviations			
Bond lengths (Å)	0.010	0.0095	
Bond angles (°)	1.484	1.503	
<i>B</i> -factor average			
Protein	78.75	152.67	
Ligand/ion	87.18	127.24	
Water	47.34	0	
<b>Ramachandran Plot Region (PROCHECK)</b>			
Most favored (%)	87.8	83.2	
Additionally allowed (%)	11.7	16.8	
Generously allowed (%)	0.5	0.0	
Disallowed (%)	0.0	0.0	
<b>PDB ID code</b>	<b>6BX7</b>	<b>6MGA</b>	

6

7

1 **Table 2**

2

3 Interface areas per EC repeat for PCDH1 ( $\delta$ 1 protocadherin), PCDH19 ( $\delta$ 2 protocadherin), PCDH $\alpha$ 7,  
4 PCDH $\beta$ 6, and PCDH $\gamma$ B7 (clustered protocadherins). The percentage of total interface areas that they  
5 cover are in brackets.

6

Protocadherin	EC1:EC4 ( $\text{\AA}^2$ )	EC2:EC3 ( $\text{\AA}^2$ )	EC3:EC2 ( $\text{\AA}^2$ )	EC4:EC1 ( $\text{\AA}^2$ )
<i>Hs</i> PCDH1 (6BX7)	486 (34%)	231 (16%)	231 (16%)	486 (34%)
<i>Dr</i> PCDH19 (5IU9)	379 (24%)	446 (28%)	452 (28%)	326 (20%)
<i>Mm</i> PCDH $\alpha$ 7 (5DZV)	501 (32%)	446 (29%)	381 (24%)	231 (15%)
<i>Mm</i> PCDH $\beta$ 6 (5DZX)	531 (24%)	660 (30%)	491 (22%)	545 (24%)
<i>Mm</i> PCDH $\gamma$ B7 (5SZP)	179 (11%)	636 (37%)	474 (28%)	414 (24%)

7