SPECIES-DEPENDENT HETEROPHILIC AND HOMOPHILIC CADHERIN INTERACTIONS IN INTESTINAL INTERMICROVILLAR LINKS

Michelle E. Gray^{*1}, Zachary R. Johnson^{*§2}, Debadrita Modak^{*2}, Matthew J. Tyska³, and Marcos Sotomayor^{#1,2}

¹Ohio State Biochemistry Program

²Department of Chemistry and Biochemistry

The Ohio State University, 484 W 12th Avenue, Columbus, OH 43210

³Department of Cell and Developmental Biology

Vanderbilt University School of Medicine, Nashville, TN 37232

^{*}Co-first authors

[§]Current address: Marshall University, Joan C. Edwards School of Medicine, Huntington, WV 25701

[#] Corresponding Author

E-mail: sotomayor.8@osu.edu

ABSTRACT

Enterocytes are specialized epithelial cells lining the luminal surface of the small intestine that build densely packed arrays of microvilli known as brush borders. These microvilli drive nutrient absorption and are arranged in a hexagonal pattern maintained by intermicrovillar links formed by two non-classical members of the cadherin superfamily of calcium-dependent cell adhesion proteins: protocadherin-24 (PCDH24, also known as CDHR2) and the mucin-like protocadherin (CDHR5). The extracellular domains of these proteins are involved in heterophilic and homophilic interactions important for intermicrovillar function, yet the structural determinants of these interactions remain unresolved. Here we present X-ray crystal structures of the PCDH24 and CDHR5 extracellular tips and analyze their species-specific features relevant for adhesive interactions. In parallel, we use binding assays to identify the PCDH24 and CDHR5 domains involved in both heterophilic and homophilic adhesion for human and mouse proteins. Our results suggest that homophilic and heterophilic interactions involving PCDH24 and CDHR5 are species dependent with unique and distinct minimal adhesive units.

1 INTRODUCTION

2 Enterocytes are specialized epithelial cells lining the luminal surface of the small intestine and are fundamental players in
3 nutrient absorption with a role in host defense (Maroux et al., 1988; Mukherjee et al., 2008; Selsted and Ouellette, 2005;
4 Shifrin and Tyska, 2012). Key to the aforementioned processes is the brush border, so named because the apical surface
5 of the enterocytes is coated by thousands of microvilli of similar length and size organized in a hexagonal arrangement
6 (Mooseker, 1985). The organization and structure of the microvilli is maintained by a network of cytoplasmic and
7 transmembrane proteins known as the intermicrovillar adhesion complex (IMAC), which includes proteins with
8 extracellular adhesive domains and cytoplasmic parts tethered to the actin cytoskeleton that forms the interior of the
9 microvilli (Crawley et al., 2016, 2014b, 2014a; Li et al., 2016; Pinette et al., 2018; Weck et al., 2016). Perturbations of
10 brush border function are often associated with disease (Bailey et al., 1989; Khubchandani et al., 2011; Vallance et al.,
11 2002; Wilson et al., 2001) and, not surprisingly, disruption of the IMAC components results in intestinal dysfunction
12 (Bitner-Glindzicz et al., 2000; Crawley et al., 2014b; Hussain et al., 2004).

13
14 The extracellular parts of the IMAC stem from a pair of cellular adhesion proteins that belong to the cadherin superfamily,
15 protocadherin-24 (PCDH24, also known as CDHR2) and mucin-like protocadherin (CDHR5) (Crawley et al., 2014b;
16 McConnell et al., 2011). These two non-classical protocadherins form intermicrovillar links essential for brush border
17 morphogenesis and function (Crawley et al., 2014a, 2014b), stabilizing the microvillar hexagonal patterns as shown by
18 freeze-etch electron microscopy and PCDH24 immuno-labeling combined with transmission electron microscopy
19 (Crawley et al., 2014b). Mutations that impair the PCDH24 and CDHR5 interaction, along with knockdowns of PCDH24
20 and CDHR5, cause a remarkable reduction in microvillar clustering *ex vivo* (Crawley et al., 2014b). A PCDH24 knockout
21 mouse model was found to be viable but body weight of mutant mice was lower than wild-type and there were defects in
22 the packing of the microvilli in the brush border (Pinette et al., 2018). Despite the important physiological role played by
23 PCDH24 and CDHR5, little is known about the molecular details of how these proteins interact to form the
24 intermicrovillar links.

25
26 PCDH24 and CDHR5 are members of a large superfamily of proteins with extracellular domains that often mediate
27 adhesion and that have contiguous and similar, but not identical, extracellular cadherin (EC) repeats. The linker regions

between the EC repeats typically bind three calcium ions essential for adhesive function (Frank and Kemler, 2002; Gumbiner, 2005; Sano et al., 1993; Takeichi, 1990). PCDH24 belongs to the Cr-2 subfamily and has 9 EC repeats, a membrane adjacent domain (MAD10), a transmembrane domain, and a C-terminal cytoplasmic domain (Bork and Patthy, 1995; De-la-Torre et al., 2018; Ge et al., 2018; Hulpiau and van Roy, 2009; McConnell et al., 2011; Pei and Grishin, 2017). CDHR5 belongs to the Cr-3 subfamily and has 4 EC repeats, a mucin-like domain (MLD), a transmembrane domain, and a C-terminal cytoplasmic domain. Some human CDHR5 isoforms lack the MLD, which is not essential for the interaction with human PCDH24 (Crawley et al., 2014b). Sequence alignments of the N-terminal repeats (EC1-3) suggest that PCDH24 and CDHR5 are similar to another pair of heterophilic interacting cadherins: cadherin-23 (CDH23) and protocadherin-15 (PCDH15), the cadherins responsible for the formation of inner-ear tip links (Elledge et al., 2010; Kazmierczak et al., 2007; Sotomayor et al., 2014, 2012, 2010). PCDH24 is most similar to CDH23, having the residues and elongated N-terminal β -strand that are predicted to favor the formation of an atypical calcium-binding site at its tip (Elledge et al., 2010; Sotomayor et al., 2010). CDHR5 is most similar to PCDH15, and features cysteine residues predicted to form a disulfide bond at its tip (Sotomayor et al., 2014, 2012). In addition, a mutation in CDHR5, R84G (residue numbering throughout the text corresponds to processed proteins, see Methods), which mimics a deafness-related mutation in PCDH15, interferes with the intermicrovillar links formed by PCDH24 and CDHR5 (Crawley et al., 2014b), suggesting that these proteins interact in a similar fashion to the heterophilic tip-link “handshake” formed by CDH23 and PCDH15 (Behrendt et al., 2012; Sotomayor et al., 2014, 2012).

To better understand how PCDH24 and CDHR5 interact to form intermicrovillar links, we used a combination of structural and bead aggregation assays to characterize the adhesive properties and mechanisms of these cadherin family members. We present the X-ray crystallographic structures of *Homo sapiens* (*hs*) PCDH24 EC1-2, *Mus musculus* (*mm*) PCDH24 EC1-3, and *hs* CDHR5 EC1-2 refined at 2.3 Å, 2.1 Å, and 1.9 Å resolution, respectively. The structures give insight into possible binding mechanisms among species, which are further probed by bead aggregation assays revealing that the human and mouse intermicrovillar cadherins do not engage in the same homophilic and heterophilic interactions. Based on these results, we suggest models for the PCDH24/CDHR5 heterophilic interaction utilized to form intermicrovillar links.

1 RESULTS

2 PCDH24 and CDHR5 tip sequences are poorly conserved across species

3 Intermicrovillar links formed by PCDH24 and CDHR5 in the enterocyte brush border are similar to inner-ear tip links
4 formed by CDH23 and PCDH15. Both types of links are essential for the development, assembly, and function of actin-
5 based structures (microvilli in the gut and stereocilia in the inner ear), and both are made of long cadherin proteins that
6 have similar cytoplasmic partners involved in interactions with the cytoskeleton and in signaling (Crawley et al., 2014a;
7 Pepermans and Petit, 2015). Sequence analyses have revealed similarities between the tips of PCDH24 and CDH23, and
8 between the tips of CDHR5 and PCDH15 (Sotomayor et al., 2014). Given that CDH23 and PCDH15 engage in a
9 heterophilic “handshake” complex involving their EC1-2 tips, it has been proposed that PCDH24 and CDHR5 might use a
10 similar binding mechanism (Crawley et al., 2014b; Sotomayor et al., 2014). To further explore this hypothesis, we
11 performed sequence analyses comparing the tips of these cadherins to each other, to classical cadherins, and across
12 species.

13
14 Interestingly, alignments of EC repeat sequences across up to 20 species for intermicrovillar and tip-link cadherins reveal
15 poor conservation for PCDH24 and CDHR5 when compared to CDH23 and PCDH15. Average percent identity is 51.6%
16 for CDH23 EC repeats and 45.5% for PCDH15 repeats, with the N- and C-terminal ends being more conserved than the
17 middle EC repeats (Choudhary et al., 2019; Jaiganesh et al., 2018). In contrast, average percent identity is 16.4% for
18 PCDH24 and 11% for CDHR5, considerably lower when compared to values for CDH23 and PCDH15. Moreover, the N-
19 and C-terminal ends of PCDH24 and CDHR5 tend to be less conserved than the middle region of these proteins (Figure
20 1A and Figure S1 and S2; Tables S1-S5). Nevertheless, multiple sequence alignments of PCDH24 and CDH23 EC1
21 repeats confirm that PCDH24 has an elongated N-terminus that should facilitate the formation of calcium-binding site 0 as
22 observed in structures of CDH23 (Figure S3A). Binding of calcium at this site in CDH23 is mediated by several acidic
23 residues, which are also present and conserved in PCDH24. Similarly, sequence alignments of CDHR5 and PCDH15
24 EC1 repeats confirm that CDHR5 has the conserved cysteine residues that form a stabilizing disulfide bond at the tip of
25 PCDH15 EC1 (Figure S3B). None of these proteins feature the tryptophan residues that mediate homophilic binding in
26 classical cadherins (Brasch et al., 2012; Parisini et al., 2007; Shapiro et al., 1995), supporting the hypothesis that PCDH24
27 and CDHR5 might form a tip-link-like handshake complex mediating adhesion.

Pairwise sequence identity computed across up to five species for the N-terminal repeats (EC1-3) of PCDH24, CDHR5, CDH23, PCDH15, and the classical cadherins E-cadherin (CDH1) and N-cadherin (CDH2), also show different trends of sequence conservation (Figure 1B and Table S6). Average percent identity in pairwise comparisons is high for CDH2 (89%), CDH23 (86%), and PCDH15 (82%), moderate for CDH1 (61%) and lowest for PCDH24 (49%) and CDHR5 (38%). This is evident when comparing sequences from the most evolutionary distant species (human and fish). For instance, *hs* and *Danio rerio* (*dr*) PCDH24 EC1-3 sequences are 39% identical, while *hs* and *dr* CDH1 EC1-3 sequences are 54% identical. Sequence differences are still large for the human and mouse PCDH24 and CDHR5 protein tips, with 75% identity for the *hs* and *mm* PCDH24 EC1-3 sequences and 66% identity for the *hs* and *mm* CDHR5 EC1-3 sequences (compared to CDH1, CDH2, CDH23, and PCDH15 EC1-3 sequences with *hs* and *mm* pairwise identities between 83% and 98%) (Figure 1B). Proteins with sequence identity as low as 30% can share similar folds (Gilson et al., 2017), yet the low PCDH24 and CDHR5 sequence conservation in multiple sequence alignments and pair-wise comparisons suggests that details of their structures and function differ across species.

Structures of human and mouse PCDH24 tips reveal N-terminal conformational variability

Previous bead aggregation assays using the *hs* PCDH24 protein revealed that its full-length extracellular domain can mediate adhesion through homophilic interactions with itself and through heterophilic interactions with *hs* CDHR5 (Crawley et al., 2014b). Members of the cadherin superfamily of proteins often engage in homophilic and heterophilic interactions using their N-terminal tips (EC1 for classical cadherins, EC1-2 for tip-link cadherins, and EC1-4 for clustered, $\delta 1$, and $\delta 2$ protocadherins) (Cooper et al., 2016; Goodman et al., 2016; Harrison et al., 2020; Modak and Sotomayor, 2019; Nagar et al., 1996; Nicoludis et al., 2016, 2015; Rubinstein et al., 2015; Schreiner and Weiner, 2010; Shapiro et al., 1995; Sotomayor et al., 2012; Takeichi, 1990). Therefore, to gain insights into the molecular basis of PCDH24-mediated adhesion and to explore the consequences of poor sequence conservation across species, we solved the crystal structures of *hs* PCDH24 EC1-2 and *mm* PCDH24 EC1-3 refined at 2.3 Å and 2.1 Å resolution, respectively (Table 1). These structures, obtained from crystals of proteins expressed in *E. coli* and refolded from inclusion bodies (see Methods), revealed both expected and unexpected architectural features.

The *hs* PCDH24 EC1-2 structure has four molecules in the asymmetric unit including residues N1 to P215 for each protomer, while the *mm* PCDH24 EC1-3 structure has one molecule in the asymmetric unit including residues N1 to T324. Both structures had good-quality electron density maps that allowed the unambiguous positioning of side chains for most of the residues (Figure S4A-B). As expected, all EC repeats in the PCDH24 structures adopt the typical seven β -strand Greek-key fold seen in extracellular domains of cadherin proteins, with β -strands labeled A to G for each repeat (Figure 2A-D). Both *hs* and *mm* PCDH24 EC2 repeats have extended F and G strands, which are linked by a disulfide bond (S187:S201) that stabilizes an unusually long FG β -hairpin near the EC1-2 linker (Figure 2A-D, Figure S4A).

The PCDH24 EC1-2 linker region is canonical in both human and mouse structures and features calcium ions bound to sites 1, 2, and 3, which are coordinated by acidic residues from the cadherin motif NTerm-XEX-DXD-D(R/Y)(D/E)-XDX-DXNDN-CTerm (Figure 2A-E and Figure S5). Interestingly, the structure of mouse PCDH24 has a non-canonical calcium-binding linker between repeats EC2 and EC3, with only two bound calcium ions and a straight conformation (Figure 2C-D, F and Figure S5). This linker lacks necessary conserved residues to bind the calcium ion that would occupy site 1 as the canonical DXE sequence is replaced with residues 172SYN174. Although the canonical DXNDN sequence motif is 213DXPDL217 at the EC2-3 linker, typical coordination of calcium at site 3 by the asparagine carboxamide oxygen is replaced by coordination through the backbone carbonyl group of a glutamate residue in our structure (E214 in mouse, Q214 in human). The last residue of the motif (typically asparagine, replaced here by L217) coordinates calcium at site 3 through its backbone carbonyl group, an interaction that is unchanged by residue variability at this position. Sequence conservation across species for the DXE/SYN and DXNDN/DXPDL motifs suggest that most species have a non-canonical linker region between repeats EC2 and EC3 with two bound calcium ions as observed in our *mm* PCDH24 EC1-3 structure. While non-canonical linker regions with two or less bound calcium ions have been shown to be flexible (Araya-Secchi et al., 2016; Jin et al., 2012; Powers et al., 2017), equilibrium and steered molecular dynamics (SMD) simulations of *mm* PCDH24 EC1-3 show that rigidity and mechanical strength are not compromised in this non-canonical linker (Figure S6).

As expected for members of the Cr-2 family of non-clustered protocadherins, the *hs* PCDH24 EC1-2 structure shows an occupied site 0 calcium-binding site at the tip of EC1, with calcium-coordinating residues similar to those observed in

structures of CDH23 (1N, 32DXDXD36, 80XDX^{TOP}; Figure 2A-B, G, Figure S7A). Unexpectedly, the *mm* PCDH24 EC1-3 structure shows an unoccupied site 0, with an extended N-terminus protruding away from EC1 as observed for classical cadherins that exchange their N-terminal strands to form adhesive bonds (Figure 2C-D, H, Figure S7B). The sequence motifs involved in calcium coordination at site 0 are conserved, suggesting that crystallization conditions, including 3M LiCl, facilitated the opening of this site. Alternatively, other differences between mouse and human sequences might be responsible for the structural divergence. Most notably, residue R86 near the FG loop and the XDX^{TOP} motif of *mm* PCDH24 EC1 seems to be incompatible with a closed N-terminal conformation and an occupied calcium-binding site 0. The open and closed conformations observed for PCDH24 EC1 will likely determine the type of homophilic or heterophilic interactions that mediate adhesion.

Crystal contacts in human and mouse PCDH24 structures suggest distinct adhesive interfaces

Crystallographic contacts observed in cadherin structures have revealed multiple physiologically relevant interfaces (Brasch et al., 2011; Cooper et al., 2016; Goodman et al., 2016; Modak and Sotomayor, 2019; Nicoludis et al., 2016). The crystallographic packing observed for the *hs* and *mm* PCDH24 structures shows various interfaces that further highlight differences across species. The asymmetric unit of the *hs* PCDH24 EC1-2 structure shows 4 molecules with two similar anti-parallel *trans* dimers (Figure S8A) arranged perpendicular to one another to form a dimer of dimers (Figure S8B). Each antiparallel dimer positions EC1 from one protomer in front of EC2 from the next protomer, slightly shifted with respect to each other and with the extended FG β -hairpins near the EC1-2 linker regions mediating the interaction head to head. The antiparallel dimers have interface areas of 979.2 Å² and 938.6 Å² for protomers A:D and B:C, respectively. These values are both larger than an empirical threshold (856 Å²) that distinguishes biologically relevant interactions from crystallographic packing artifacts (Ponstingl et al., 2000). Two other interfaces are small and unlikely to be biologically relevant (Figure S8C-D). Previous data has shown that *hs* PCDH24 mediates homophilic *trans* adhesion when *hs* CDHR5 is not present (Crawley et al., 2014b). It is possible that the antiparallel EC1-2 interface facilitates homophilic adhesion mediated by *hs* PCDH24, yet our SMD simulations suggest that this interface is weak (Figure S9).

In contrast to the crystal contacts observed in the *hs* PCDH24 EC1-2 structure, the crystal packing of *mm* PCDH24 EC1-3 (space group *P*321) results in multiple large interfaces generated from rotations of a single molecule in the asymmetric

unit. A parallel trimer is present around the axis of symmetry, with the EC1 N-termini from each protomer protruding away and inserting into pockets of adjacent EC1 repeats in a clockwise fashion, thereby interlocking the EC1 domains (Figure S10A-B). This is possible because of the open conformation observed for the mouse N-terminus of EC1 without a bound calcium ion at site 0, and it is reminiscent of how the EC1 N-terminus of classical cadherins is exchanged resulting in the insertion of W2 into an hydrophobic binding pocket of the partner molecule (Figure S7C-D) (Parisini et al., 2007; Shapiro et al., 1995). Interestingly, a glycosylation site is predicted at N9, which might interfere with or regulate the formation of interlocking EC1 repeats, as has been observed for other cadherins (Brasch et al., 2011).

The interlocking of mouse EC1 repeats in the trimer facilitates the formation of two parallel *cis* interfaces between EC1-3 protomers involving the FG β -hairpin interacting with the neighboring EC2-3 linker. These *cis* interfaces are large with an interface area of 1352.6 Å² between two monomers (Figure S10A). In addition, a fully overlapped antiparallel *trans* interface is observed between EC1-3 protomers with an area of 1221.2 Å² (Figure S10C). Sugars at a second predicted glycosylation site located in the center of the *cis* trimer (p.N161), and at a third predicted glycosylation site near the *trans* interface between EC1 and EC3 (p.N280), may interfere with or regulate the formation of these interfaces. Together, the *cis* trimer and *trans* dimer could form a glycosylation-modulated interlocking arrangement of PCDH24 molecules that is different from the *trans* interaction in the *hs* PCDH24 EC1-2 structure. Sequence conservation mapped to the *mm* EC1-3 structure reveals that only core residues are highly conserved, highlighting poor conservation for surface exposed residues potentially involved in adhesive interactions (Figure S11). Motivated by the sequence and structural differences observed for the human and mouse PCDH24 tips, we used bead aggregation assays to determine the minimum adhesive unit of PCDH24 homophilic binding in both species.

Human, but not mouse PCDH24 mediates homophilic adhesion

To determine which EC repeats are essential for *hs* and *mm* PCDH24 homophilic adhesion we created a library of constructs encoding for the full-length extracellular domains (EC1-MAD10) and various truncations (EC1-7, EC1-6, EC1-5, EC1-4, EC1-3, EC1-2, and EC1), all fused to a C-terminal Fc tag. These protein fragments were produced in HEK293T cells and used for aggregation assays with Protein G magnetic beads (see Methods).

As expected, bead aggregation experiments using *hs* PCDH24 EC1-MAD10Fc showed large calcium-dependent aggregates (Figure 3A, L-M, Figure S15A, I). Similar large calcium-dependent aggregates were seen for *hs* PCDH24 EC1-7Fc through EC1-3Fc (Figure 3B-F, L-M, Figure S15B-F). Further truncations indicate that smaller *hs* PCDH24 fragments, including just EC1, are capable of facilitating the formation of small bead aggregates that increase in size with rocking but never reach the larger sizes observed when using fragments with 3 or more EC repeats (Figure 3G-H, K-M, Figure S15G-H). Controls performed using EDTA showed that constructs are unable to aggregate when calcium is depleted, except when using EC1Fc, possibly stabilized by the adjacent Fc-tag (Figure 3I-M, Figure S15I). These results suggest that *hs* PCDH24 mediates calcium-dependent homophilic adhesion, with at least three distinct modes of adhesion, two calcium-dependent mediated by *hs* PCDH24 EC1-3 and PCDH24 EC1-2, and another one calcium-independent mediated by *hs* PCDH24 EC1 alone. It is possible that each of these modes uses a different interface, and that the antiparallel interface observed in our *hs* PCDH24 EC1-2 structure is a low-affinity intermediate state that facilitates further interdigitation for the adhesive mode that drives the formation of larger bead aggregates.

Given the low sequence identity between the human and mouse N-terminal EC repeats of PCDH24 when compared to CDH23, and the structural differences observed between *hs* PCDH24 EC1-2 and *mm* PCDH24 EC1-3, bead aggregation assays were also performed with the full-length extracellular domain of *mm* PCDH24. Bead aggregation was not observed for *mm* PCDH24 EC1-MAD10Fc (Figure 4A-B, E, G). These results suggest that the large crystallographic antiparallel *trans* interface formed by PCDH24 EC1-3 (Figure S10C) does not mediate bead aggregation under the conditions tested and reveal that PCDH24 mediated homophilic adhesion is species-dependent.

Mouse, but not human CDHR5 mediates homophilic adhesion

The full-length extracellular domain of *hs* CDHR5 was previously shown to not mediate homophilic adhesion (Crawley et al., 2014b). The low sequence identity between the human and the mouse CDHR5, especially at their N-termini, and the species-dependent homophilic adhesive behavior of PCDH24, prompted us to test if *mm* CDHR5 could mediate adhesion. Full-length cadherin extracellular domains (without MLDs) of both human and mouse CDHR5 were used for bead aggregation assays. As expected, *hs* CDHR5 EC1-4Fc did not mediate bead aggregation in the presence and the absence

of calcium. However, *mm* CDHR5 EC1-4Fc did mediate bead aggregation in the presence of calcium (Figure 4C-D, F-G).

These results suggest that, like PCDH24, CDHR5 homophilic adhesiveness is species-dependent.

To determine the minimal adhesive unit of *mm* CDHR5, a library of truncated cadherin fragments was generated and used for bead aggregation assays. Calcium-dependent bead aggregation was observed when using beads coated with *mm* CDHR5 EC1-4Fc and EC1-3Fc, but not when using EC1-2Fc and EC1Fc (Figure 5A-D, G-I, Figure S16A-E). Hence, *mm* CDHR5 EC1-3 is sufficient to mediate homophilic adhesion.

To provide insight into the interface used by *mm* CDHR5 to mediate homophilic adhesion, we tested mutations that alter a site similar to that mutated in *hs* CDHR5 (R84G) preventing heterophilic binding to *hs* PCDH24. Bead aggregation assays were performed with both *mm* CDHR5 EC1-4Fc R82G and E84G, and both mutants abolished bead aggregation of *mm* CDHR5 EC1-4Fc in the presence of calcium (Figure 5E-F, H-I). These results suggest that the interface used for mediating heterophilic adhesion between PCDH24 and CDHR5 and the interface used for mediating homophilic adhesion by *mm* CDHR5 might be similar.

Structure of human CDHR5 EC1-2 reveals a PCDH15-like fold and suggests adhesive interfaces

To gain insight into CDHR5's adhesive function we solved the crystal structure of *hs* CDHR5 EC1-2 refined at 1.9 Å resolution (Table 1). This structure, obtained using protein expressed in *E. coli* and refolded from inclusion bodies (see Methods), had one molecule in the asymmetric unit including residues Q1 to L207, with clear electron density for residues A2 to A205 (Figure 2I, J and Figure S4C). Like other cadherins, the structure of *hs* CDHR5 EC1-2 shows straight consecutive EC repeats, each with a typical seven β-strand Greek-key fold (Figure 2I, J). The linker region between the repeats has the canonical acidic residues that coordinate calcium and three calcium ions bound to it (Figure 2K and Figure S12). Similar to PCDH15, a member of the Cr-3 subfamily of non-clustered protocadherins, *hs* CDHR5 EC1 has a disulfide bridge between cysteine residues C5 and C73 on strands A and F respectively (Figure 2L, Figure S4C, S7E). This disulfide bond keeps the N-terminal strand tucked to the protomer in a closed conformation, reminiscent of how calcium at site 0 facilitates closing of the N-terminal strand in EC1 repeats of CDH23 and *hs* PCDH24.

Next, we analyzed crystallographic contacts in the *hs* CDHR5 EC1-2 structure (Figure S13), which could lead to an understanding of how mouse, but not human CDHR5 mediates homophilic adhesion and how the heterophilic complex with PCDH24 is formed. The two largest contacts within the *hs* CDHR5 EC1-2 crystal structure are interfaces in antiparallel arrangements. The largest, at 1433.9 Å², is formed by interactions between the EC1 repeat of one protomer and EC2 of the adjacent one, with a slight shift in the register of the antiparallel dimeric interface (Figure S13A). There are up to three predicted N-glycosylation sites at this interface in the sequences for the human (p.N19 and p.N56) and the mouse (p.N17, p.N56, and p.N107) proteins. The next largest contact involves interactions between antiparallel EC2 repeats facing each other (Figure S13B), with an interface area of 530.2 Å², small relative to likely physiologically relevant interfaces (>856 Å²), but comparable to interface area per EC repeat reported for physiologically relevant contacts in protocadherins (Cooper et al., 2016; Modak and Sotomayor, 2019; Nicoludis et al., 2019). In this arrangement, the EC3 repeat (not present in the structure), would interact with EC1, thus increasing the total contact area. None of the predicted N-glycosylation sites for the human or mouse sequences are at the EC2 interface of this *trans* dimer. An X-shaped crystallographic contact is also present (Figure S13C), with a small interface area (364.7 Å²) involving EC1 to EC1 interactions. Additional smaller interfaces are present but are unlikely to be physiologically relevant given their atypical arrangement and small interface areas of 347.9 Å², 132.7 Å², and 114.0 Å², respectively (Figure S13D-F).

Although, *hs* CDHR5 does not mediate adhesion in bead aggregation assays, some of the contacts discussed above can serve as templates to predict the interfaces that could be used by *mm* CDHR5 to mediate homophilic adhesion. The largest, antiparallel EC1-2 interface might be impaired by glycosylation at a site predicted to exist in the mouse sequence (N107), but not in the human protein (R107). In addition, bead aggregation assays suggest that repeat EC3 is needed for adhesive interactions, suggesting that this EC1-2 interface is not driving CDHR5 homophilic adhesion. In contrast, the antiparallel EC1-3 interface mediated by EC2 interactions is compatible with results from bead aggregation assays indicating that the first three N-terminal repeats are sufficient for *trans* adhesiveness, but details of the arrangement are likely to be different due to the absence of EC3 in our structure and the low sequence conservation of surface residues (Figure S14). These contacts may also guide the search for interaction modes that facilitate the formation of PCDH24 and CDHR5 heterophilic complexes.

Human and mouse heterophilic PCDH24 and CDHR5 bonds are distinct

Previous bead aggregation assays showed that full-length extracellular domains of human PCDH24 and CDHR5 mediate heterophilic adhesion (Crawley et al., 2014b). We used our library of Fc-tagged PCDH24 and CDHR5 cadherin fragments to test for heterophilic adhesiveness for both human and mouse proteins. To distinguish PCDH24 from CDHR5 coated beads, we used green and red fluorescent beads, respectively. We confirmed that human PCDH24 and CDHR5 cadherin extracellular domains mediate calcium-dependent heterophilic adhesion, and found that the mouse proteins also facilitate calcium-dependent bead aggregation (Figure 6A-B, I, J, K and Figure S17A-B), suggesting that the heterophilic intermicrovillar cadherin bond driving intestinal brush border assembly is species independent.

Next, we used truncated protein fragments and found that repeats EC1-4Fc of PCDH24 and CDHR5 from both species are sufficient to drive heterophilic bead aggregation (Figure 6A-B, I, J, K and Figure S17A-B). Further bead aggregation assays focused on determining the minimum adhesive units for heterophilic adhesion. We carried out heterophilic binding assays using PCDH24 EC1-4Fc mixed with truncated fragments of CDHR5 and vice versa for both human and mouse proteins (Figure 6C-F, I, J, K and Figure S17C-F). The minimal adhesive units for heterophilic binding when using human proteins were PCDH24 EC1-2Fc and CDHR5 EC1Fc, whereas mouse proteins required PCDH24 EC1-2Fc and CDHR5 EC1-2Fc.

Additional experiments using only the required EC repeats of PCDH24 and CDHR5 to test heterophilic adhesiveness for both human and mouse proteins confirmed the results obtained with the truncation series (Figure 6G, K and Figure S17G, I). Binding assays in the presence of 2 mM EDTA abolished aggregation indicating that heterophilic adhesion mediated by PCDH24 and CDHR5 is calcium-dependent (Figure 6H, K and Figure S17H, J). The low sequence similarity of CDHR5 and PCDH24 across different species may explain the differences in their binding mechanisms, as demonstrated by the distinct minimal adhesive units found for the human and mouse heterophilic complexes. Remarkably, while homophilic adhesive behavior is not conserved between human and mouse, the heterophilic bond is robust with slight differences in underlying molecular mechanisms.

DISCUSSION

Our sequence analyses, structures, and bead aggregation assays suggest that there are various modes of *trans* homophilic and heterophilic interactions involving the extracellular domains of PCDH24 and CDHR5 (Figure 7). We have confirmed that human PCDH24 can form homophilic bonds (Crawley et al., 2014b), and found three potentially distinct modes of *trans* adhesion in bead aggregations assays. The first mode involves weak (small bead aggregates) calcium-independent interactions mediated by EC1. This adhesive mode is not strong enough to drive calcium-dependent PCDH24 adhesion when using larger parts of its extracellular domain, and therefore might be unphysiological or relevant only when a pre-formed bond is exposed to calcium chelators. The physiological PCDH24 *trans* homophilic bond might be initiated by weak interactions involving EC1 and EC2 repeats (small bead aggregates), with further interdigitation for stronger interactions (larger bead aggregates) mediated by at least EC1 to EC3 repeats.

Intriguingly, our bead aggregation assays indicate that mouse PCDH24 does not mediate *trans* homophilic adhesion, highlighting a species-dependent behavior that is consistent with poor sequence and structural conservation for the mouse and human proteins. In contrast, experiments with CDHR5 reveal that the mouse protein can mediate *trans* homophilic adhesion, but the human protein cannot. These results are consistent with L929 cell aggregation assays that showed homophilic adhesion mediated by the rat CDHR5 protein (Goldberg et al., 2002, 2000). The physiological relevance of the *trans* homophilic bonds for human PCDH24 and mouse CDHR5 remains to be determined. Mice lacking PCDH24 expression in the intestinal track are viable, but intestinal tissue exhibits various structural defects leading to functional impairment of intestine function and loss of body weight (Pinette et al., 2018). Homophilic CDHR5 bonds might be partially compensating for the lack of heterophilic intermicrovillar links in these mice.

Importantly, adhesion mediated by PCDH24 and CDHR5 heterophilic bonds is species independent in our bead aggregation assays with mouse and human proteins. These results also demonstrate that lack of homophilic adhesion for mouse PCDH24 and for human CDHR5 is not caused by issues with bead coating or with protein misfolding and degradation in our assays. Details of the heterophilic interaction between PCDH24 and CDHR5, however, might differ from species to species, as we found that the minimal adhesive units are PCDH24 EC1-2Fc / CDHR5 EC1Fc for human proteins and PCDH24 EC1-2Fc / CDHR5 EC1-2Fc for mouse proteins.

1

2 Our structures of the human PCDH24 EC1-2, mouse PCDH24 EC1-3, and human CDHR5 EC1-2 protein fragments
3 confirm the expected structural similarities between the PCDH24 and CDH23 tips, and between the PCDH15 and CDHR5
4 tips. These structures also support binding mechanisms that are distinct from those used by classical cadherins, clustered
5 protocadherins, and non-clustered δ protocadherins (Brasch et al., 2019, 2012; Cooper et al., 2016; Goodman et al., 2016;
6 Harrison et al., 2020; Lefebvre, 2017; Modak and Sotomayor, 2019; Nicoludis et al., 2016, 2015; Rubinstein et al., 2015).
7 While crystal contacts in our structures suggest possible interfaces that might facilitate the modes of adhesion discovered
8 through bead aggregation assays, further work will be required to elucidate and validate structural models of *trans*
9 homophilic and heterophilic complexes, to understand the relevance of an *in crystallo cis* complex observed here for
10 mouse PCDH24, to clarify the structural role played by glycosylation *in vivo* (Krahn et al., 2010), and to determine which
11 complexes form intermicrovillar links in enterocytes.

12

13 Intermicrovillar links observed in human CACO-2_{BBE} cells and in native mouse tissue are 46.8 ± 8.9 nm and 49.9 ± 8.8
14 nm in length, respectively (Crawley et al., 2014b). Interestingly, lengths range from 30 to 80 nm for both human and
15 mouse links, perhaps reflecting the challenges of measuring link sizes from two-dimensional electron-microscopy images
16 of cells, with some additional uncertainty coming from difficult-to-define link edges. Nevertheless, this range of lengths
17 hints at multiple possibilities for the architecture of the complex that are consistent with the modes of adhesion we found
18 using bead aggregation assays. Links made of two PCDH24 molecules interacting in *trans* and tip-to-tip could reach
19 lengths of up to ~90 nm (9 EC repeats + MAD10 per molecule; 4.5 nm long each), while links made of the shortest
20 CDHR5 isoforms with just 4 EC repeats interacting tip-to-tip could be 36 nm in length or even shorter if there is any
21 overlap at the junction or bending at any of the EC linkers. Heterophilic links formed by a tip-to-tip *trans* interaction of
22 the full-length PCDH24 ectodomain and the shortest CDHR5 isoform should be ~63 nm or shorter. The average value
23 observed *ex vivo* and *in vivo* (~48 nm) indicates that some overlap at the junction may exist, and that most of the links are
24 likely formed by the heterophilic complex. The minimal adhesive units we found for mouse and human PCDH24 and
25 CDHR5 homophilic and heterophilic *trans* complexes are consistent with these estimates for length ranges, assuming that
26 up to 3 EC repeats might be overlapping to form robust bonds that lead to bead aggregation *in vitro*, and adhesion *in vivo*.

The functional role of PCDH24 and CDHR5 in non-intestinal tissues and cells is unclear. These proteins have also been found in the liver and kidney (Goldberg et al., 2000; Okazaki et al., 2002) and have been implicated in contact inhibition of cell proliferation (Okazaki et al., 2002; Ose et al., 2009), morphogenesis (Behrendt et al., 2012; Goldberg et al., 2000; Krahn et al., 2010), colorectal carcinogenesis (Bujko et al., 2015; Losi and Grande, 2014), and gallstone disease (Chuang et al., 2011). Exploring how different organs in different species express and use PCDH24 and CDHR5 proteins, which have poorly conserved sequences, might provide with a unique opportunity to reveal how evolutionary events structurally encode adhesive and signaling functions in various physiological contexts (Dickinson et al., 2011). It is tempting to speculate that differences in gut proteases' specificity across species, as well as diverse diets and gut microbiota, put divergent selective pressures driving the molecular evolution of intermicrovillar links. Similarly, other selective pressures relevant to specific organs might be driving the evolution of heterophilic and homophilic PCDH24 and CDHR5 adhesive complexes.

Interestingly, both brush border enterocytes and inner-ear sensory receptor cells use heterophilic cadherin complexes in actin-based structures relevant for their specialized functions. These heterophilic cadherin complexes link membranes within cells, as opposed to mostly homophilic cadherin complexes that mediate adhesion between membranes of adjacent cells. Heterophilic complexes, with a pair of distinct cytoplasmic domains, might be more functionally versatile in signaling, as the number and type of cytoplasmic partners can be independently diversified for each member of the complex. The strength and dynamics of cadherin heterophilic bonds, which are present in systems subjected to mechanical stimuli, might also differ from and provide specific advantages over weak homophilic cadherin complexes that cluster to provide strong adhesiveness. Whether PCDH24 and CDHR5 engage in *cis* interactions that can lead to more mechanically robust intermicrovillar links, as observed for tip links (Choudhary et al., 2019; Dionne et al., 2018; Ge et al., 2018), or to the formation of larger adhesive complexes in non-microvillar membranes, as observed for classical cadherins and clustered protocadherins (Brasch et al., 2019, 2012; Schreiner and Weiner, 2010), remains to be explored.

METHODS

Cloning, expression, and purification of PCDH24 and CDHR5 fragments for crystallography

DNA encoding for *hs* PCDH24 EC1-2, *mm* PCDH24 EC1-3 and *hs* CDHR5 EC1-2 were sub cloned into *NdeI* and *XhoI* sites of the pET21a vector including a fused C-terminal hexa-histidine tag. All DNA constructs were sequence verified. These constructs were used to transform BL21(DE3) pLysS cells (Stratagene). The PCDH24 transformants were cultured in TB and induced at OD₆₀₀ ~ 0.6 with 1 mM IPTG at 37°C overnight. The *hs* CDHR5 EC1-2 transformant was cultured in TB and induced at an OD₆₀₀ ~ 0.6 with 1 mM IPTG at 30°C overnight. Cells were lysed by sonication in denaturing buffer (20 mM TrisHCl [pH7.5], 6 M guanidine hydrochloride [GuHCl], 10 mM CaCl₂ and 20 mM imidazole). The cleared lysates were loaded onto Ni-Sepharose beads and nutated for ~1 h (GE Healthcare). Beads were washed twice with denaturing buffer and hexa-histidine-tagged proteins were eluted with the same buffer supplemented with 500 mM imidazole. After purification, 2 mM DTT was added to the protein before refolding. The *hs* PCDH24 EC1-2 protein was refolded in six dialysis steps using a buffer with 20 mM TrisHCl [pH 8.0], 5 mM CaCl₂ and decreasing amounts of GuHCl (6 M, 3 M, 2 M, 1 M, 0.5 M and 0 M). In steps four, five and six the refolding buffer included 400 mM arginine. Steps three through six also included 150 mM NaCl. The *mm* PCDH24 EC1-3 and *hs* CDHR5 EC1-2 proteins were refolded overnight using a buffer with 400 mM Arginine, 20 mM TrisHCl [pH 8.0], 150 mM NaCl, and 5 mM CaCl₂. All dialysis steps were performed using MWCO 2000 membranes and used refolding buffers for 12 hours at 4°C. Refolded protein was further purified on a Superdex200 16/600 column (GE Healthcare) in 20 mM TrisHCl [pH 8.0], 150 mM NaCl and 2 mM CaCl₂, and concentrated by ultrafiltration to ~5 mg/ml for crystallization (Vivaspin 10 KDa).

Crystallization, data collection, and structure determination

Crystals were grown by vapor diffusion at 4°C by mixing 1 µL of protein and 0.5 µL of reservoir solution including (0.2 M MgCl, 0.1 M HEPES [pH 7.3], 10% PEG 4000) for *hs* PCDH24 EC1-2, (3.0 M LiCl, 0.1 M HEPES [pH 7.5]) for *mm* PCDH24 EC1-3 and (0.1 M Na Acetate pH 4.4, 0.7 M CaCl₂) for *hs* CDHR5 EC1-2. Crystals were cryo-protected in reservoir solution plus a cryoprotectant (25% PEG 400 for PCDH24 and 25% glycerol for *hs* CDHR5 EC1-2) and cryo-cooled in liquid N₂. X-ray diffraction datasets were collected as indicated in Table 1 and processed with HKL2000 (Otwinowski and Minor, 1997).

All structures were solved by molecular replacement using PHASER (McCoy et al., 2007). The first two N-terminal repeats of CDH23 EC1-2 (PDB: 4AQE) (Sotomayor et al., 2012) were used as an initial model for the *mm* PCDH24 EC1-

3 structure. The *mm* PCDH24 EC3 repeat was built using BUCCANEER (Cowtan, 2007, 2006). The *hs* PCDH24 EC1-2 structure was determined by using the first two N-terminal EC repeats of the *mm* PCDH24 EC1-3 structure (PDB: 5CYX). For the *hs* CDHR5 EC1-2 structure we used a model of CDH23 EC22-24 (PDB: 5UZ8) (Jaiganesh et al., 2018) created using RaptorX (Källberg et al., 2012). Model building was done with COOT (Emsley et al., 2010) and refinement was performed with REFMAC5 (Murshudov et al., 2011). Restrained TLS refinement was used for the PCDH24 structures. Data collection and refinement statistics are provided in Table 1.

Cloning of Fc-fusions proteins for bead aggregation assays

DNA constructs encoding C-terminal truncations of *hs* and *mm* PCDH24 as well as of *hs* and *mm* CDHR5, all fused to a C-terminal Fc-tag, were prepared for bead aggregation assays. Truncations were made at DXNDN or similar ends of each EC repeat (Figure S1 and S2). Truncations were cloned into a modified CMV vector (Jontes laboratory, OSU) using *XhoI* and *KpnI* (*hs* PCDH24, *mm* PCDH24), *NheI* and *XhoI* (*hs* CDHR5), and *XhoI* and *BamHI* (*mm* CDHR5) cutsites. The sequence for the linker between the *hs* CDHR5 protein fragments and their Fc tags is LELKLRILQSTVPRARDPPV, while the sequence for the linker between the *mm* CDHR5 protein fragments and their Fc tags is TDPPV. Mutations were generated using the QuikChange Lightning kit (Agilent). All DNA constructs were sequence verified.

Bead aggregation assays

Bead aggregation assays were performed as previously described (Cooper et al., 2016; Crawley et al., 2014b; Emond et al., 2011; Emond and Jontes, 2014; Modak and Sotomayor, 2019). The PCDH24Fc and CDHR5Fc fusion constructs were transfected into HEK293T cells via calcium-phosphate transfection (Barry et al., 2014; Jiang and Chen, 2006; Kwon and Firestein, 2013). A solution of 10 µg of DNA and 250 mM CaCl₂ was added dropwise to 2X HBS with mild vortexing. The solution was immediately added drop-wise to a 100 mm dish of HEK293T cells cultured in DMEM with FBS, L-glut, and PenStrep. For each construct, two plates of cells were transfected. The following day, cells were rinsed twice with 1X PBS and allowed to grow in serum-free media for two days prior to collection. The collected media contained the secreted Fc fusion proteins and was concentrated (Amicon 10 kD, 30 kD concentrators) to a volume of 500 µL before being incubated with 1.5 µL of Protein G Dynabeads (Invitrogen). The beads were incubated with rotation for 2 hr at 4°C. Incubated beads were washed with binding buffer (50 mM Tris [pH 7.5], 100 mM NaCl, 10 mM KCl, and 0.2% BSA) and

split into two tubes, either with 2 mM CaCl₂ or 2 mM EDTA. Beads were then allowed to aggregate on a glass depression slide in a humidified chamber for 1 h, then rocked for 1 to 2 min at 8 oscillations/min. Images were taken at various time points using a Nikon Eclipse Ti microscope with a 10X objective as detailed in Table S7. Sizes of bead aggregates were quantified with ImageJ as described previously (Emond et al., 2011; Emond and Jontes, 2014). Images were thresholded to exclude background and the area of the detected particles was measured in pixels. Average aggregate size ($\pm$ SEM) was plotted in Origin Pro. Number of independent biological replicates are listed in Table S7.

Fluorescent bead aggregation assays

Fluorescently labeled protein A Dynabeads were produced for the heterophilic bead aggregation assays. Protein A (5 mg; Thermo Fisher Pierce) was resuspended in 1X PBS to obtain a final concentration of 2 mg/mL. Using the Thermo Fisher Dylight Antibody Labeling kits, 1 mg of Protein A was labeled with Dylight 488 to produce 0.5 moles of dye/1 mole of Protein A. Labeled Protein A (200 μ g) was then conjugated to Dynabeads M-280 Tosylactivated (Invitrogen) following the provided protocol to produce 10 mg of beads at a final concentration of 20 mg/mL. A similar protocol was used to obtain Dylight 594 conjugated Protein A Dynabeads. Conjugation to the beads was confirmed by observing the beads in depression slides under 10X magnification with the GFP filter cube (Nikon) for the Dylight 488 fluorophore and the Texas Red filter cube (Nikon) for the Dylight 594 fluorophore. Beads of both colors were then mixed on a depression slide and no overlap of fluorescence was observed. Further confirmation that the beads were functional was done by performing a bead aggregation assay with a positive control (*hs* PCDH24 EC1-4).

Fluorescent bead aggregation assays were performed as done with non-fluorescent beads but with noted exceptions. Concentrated media was added to 2.2 μ L of fluorescently labeled beads to achieve the same concentration as the Protein G Dynabeads on the slides. Dylight 488 conjugated beads were used for all PCDH24Fc fusions and Dylight 594 conjugated beads were used for all CDHR5Fc fusions. Prior to addition of beads to the CaCl₂ or EDTA tubes, 160 μ L of Dylight 488 beads containing the PCDH24 constructs were mixed with 160 μ L Dylight 594 beads containing CDHR5 constructs, then 150 μ L of the mixture was added to the CaCl₂ or EDTA tubes.

Western Blots

Western blots were performed for Fc-tagged proteins to confirm expression and secretion of the protein. Protein samples were mixed with 4X SDS with β -mercaptoethanol, boiled for 5 min, and loaded onto SDS-PAGE gels (BioRad) for electrophoresis. Proteins were electroblotted onto PVDF membrane (GE Healthcare), blocked with 5% nonfat milk in TBS with 0.1% Tween-20 for 1 h, and incubated overnight at 4°C with goat anti-human IgG (1:200; Jackson ImmunoResearch Laboratories, Catalog-109-025-003, Lot-117103, 135223). After washing with 1X TBS, the membranes were incubated with mouse anti-goat HRP-conjugated secondary antibody (1:5000; Santa Cruz Biotechnology, Catalog-sc-2354, Lot-A2017, J1718) for 1 h at room temperature. Membranes were washed in 1X TBS and developed using the ECL Select Western Blot Detection kit (GE Healthcare) for chemiluminescent detection (Omega Lum G). Western blots were performed on samples obtained immediately after media collection and concentration as well as protein samples obtained from the Protein G/A beads following aggregation assays.

Sequence alignments

Sequence alignments were performed on individual EC repeats for a variety of species of PCDH24 (Table S2) and CDHR5 (Table S3) in MUSCLE (Edgar, 2004). Sequences that covered the complete extracellular domain were selected to include a variety of species of each protein. Sequence conservation per EC repeat in percent identity was obtained from the MUSCLE alignments in Geneious (Kearse et al., 2012) (Table S1). All EC repeats of both human and mouse PCDH24 (Figure S5) and CDHR5 (Figure S12) were aligned to one another as well as described above. Conservation of alignments were visualized in JalView (Waterhouse et al., 2009) using percent identity (Henikoff and Henikoff, 1992) with a conservation cutoff of 40%. Additionally, sequence alignments of individual EC repeats for a variety of species of PCDH15 (Choudhary et al., 2019) (Table S4) and CDH23 (Jaiganesh et al., 2018) (Table S5) were performed in MUSCLE, and sequence conservation per EC repeat was obtained in Geneious (Table S1). A separate set of alignments, which include up to 5 species for the N-terminal repeats (EC1-3) of PCDH24, CDHR5, CDH23, PCDH15, CDH1, and CDH2 (Table S6), were performed in MUSCLE. The Sequence Identity and Similarity Server (SIAS) was used to determine identity between pairs of different species for each protein (Figure 1B). A broader conservation analysis of PCDH24 EC1-3 and CDHR5 EC1-2 was done in Consurf (Ashkenazy et al., 2016, 2010; Celniker et al., 2013; Glaser et al., 2003; Landau et al., 2005) using additional sequences to include a larger variety of species (Table S8 for PCDH24 and Table S9 for CDHR5). The values obtained from Consurf (1 = lowest conservation, 9 = highest conservation) were

matched to the corresponding residues in the structures and used to make a heat map of conservation. N-glycosylation and O-glycosylation sites for the extracellular domains of PCDH24 and CDHR5 were predicted using NetNGlyc 1.0 and NetOgly 4.0 (Steentoft et al., 2013) and highlighted in the multiple sequence alignment. All positives for O-glycosylation were labeled, but predictions for N-glycosylation were only labeled if the prediction received a 9/9 prediction score from the server.

Simulated systems

The psfgen, solvate, and autoionize VMD (Humphrey et al., 1996) plugins were used to build all systems (Table S10), adding hydrogen atoms to the protein structure and crystallographic water molecules. For the *mm* PCDH24 EC1-3 systems, the missing loop in EC1 (residues 32 to 37) was built in COOT (Emsley et al., 2010) using the human PCDH24 EC1 structure as a template. Non-native N-terminal residues were removed. Residues D, E, K, and R were considered to be charged and neutral histidine protonation states were determined on favorable hydrogen bond formation. The proteins were placed in additional water and randomly placed ions to solvate and ionize the systems at 150 mM NaCl. For SMD simulations, protein molecules were aligned to the *x*-axis. Additional information for systems is presented in (Table S10).

Simulations Parameters

Equilibrium and SMD simulations were performed using NAMD 2.11 (Phillips et al., 2005), the CHARMM36 forcefield for proteins with the CMAP correction, and the TIP3P model for water (Huang and MacKerell, 2014). Van der Waals interactions were computed with a cutoff at 12 Å and using a switching function starting at 10 Å. Computation of long-range electrostatic forces was done with the Particle Mesh Ewald method with a cutoff of 12 Å and a grid point density of $>1 \text{ Å}^{-3}$. A 2 fs uniform integration time was used with SHAKE. When indicated, constant temperature ($T = 300\text{K}$) was enforced using Langevin dynamics with a dampening coefficient of 0.1 ps^{-1} unless otherwise noted. The hybrid Nose-Hoover Langevin piston method was used to maintain constant pressure (NpT) at 1 atm with a 200 fs decay period and a 100 fs dampening time constant. In simulation S3a (Table S10), the *mm* PCDH24 EC1-3 protein had harmonic constraints ($k = 1 \text{ kcal mol}^{-1} \text{ Å}^{-2}$) applied to the C_{α} atoms of residues S110, V123, and L182 in EC2 to avoid rotations leading to interactions between periodic images. Constant velocity stretching simulations (S1b-d and S3b-d, Table S10) were carried out using the SMD method and the NAMD Tcl interface (Franz et al., 2020; Grubmüller, 2005; Israelewitz et

al., 2001; Sotomayor and Schulten, 2007). Virtual springs ($k_s = 1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) were attached to C α atoms of C-terminal ends for unbinding simulations (S1b-d) and to the center of mass of EC1 and EC3 for unfolding simulations (S3b-d). The free ends of the stretching springs were moved in opposite directions, away from the protein(s), at a constant velocity. Applied forces were computed using the extension of the virtual springs. Maximum force peaks and their averages were computed from 50 ps running averages.

ACCESSION NUMBERS

Coordinates for *hs* PCDH24 EC1-2, *mm* PCDH24 EC1-3 and *hs* CDHR5 EC1-2 have been deposited in the Protein Data Bank with entry codes 5CZR, 5CYX, and 6OAE, respectively.

ACKNOWLEDGEMENTS

We thank members of the Sotomayor laboratory for assistance and discussions. This work was supported in part by the Ohio State University and NIH (NIDDK R01 DK095811 to M.J.T.). Use of APS NE-CAT beamlines was supported by NIH (P41 GM103403 & S10 RR029205) and the Department of Energy (DE-AC02-06CH11357) through grants GUP 49774 and 59251. Use of TACC-Stampede and PSC-Bridges supercomputers was supported by the National Science Foundation through XSEDE (XRAC MCB140226). Use of the OSC-Owens supercomputer was supported by grant PAS1037. DM was a Pelotonia fellow and MS was an Alfred P. Sloan fellow (FR-2015-65794).

AUTHOR CONTRIBUTIONS

M.E.G. did sequence alignments, solved the structure for *hs* CDHR5 EC1-2, did cloning of *hs* and *mm* PCDH24Fc constructs, performed homophilic binding assays of PCDH24, prepared the fluorescence magnetic beads for heterophilic bead aggregation assays, and did simulations. Z.R.J. did cloning, protein expression and protein purification, crystallization trials, and solved structures for *hs* PCDH24 EC1-2, *mm* PCDH24 EC1-3, and *hs* CDHR5 EC1-2. D.M. did cloning of *hs* and *mm* CDHR5Fc constructs, carried out homophilic bead aggregation assays of CDHR5 and did all heterophilic bead aggregation assays. M.J.T. provided guidance, reagents, and secured funding. M.S. trained M.E.G., Z.R.J., and D.M. and assisted with crystal fishing, cryo-cooling, solving crystal structures, and simulations. M.S.

supervised work and assisted with data analysis. M.E.G., D.M., and M.S. prepared figures and wrote the manuscript with feedback from M.J.T. and Z.R.J.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

- Araya-Secchi R, Neel BL, Sotomayor M. 2016. An elastic element in the protocadherin-15 tip link of the inner ear. *Nat Commun* **7**:13458. doi:10.1038/ncomms13458
- Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. doi:10.1093/nar/gkw408
- Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. 2010. ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* **38**:529–533. doi:10.1093/nar/gkq399
- Bailey DS, Freedman AR, Price SC, Chescoe D, Ciclitira PJ. 1989. Early biochemical responses of the small intestine of coeliac patients to wheat gluten. *Gut* **30**:78–85. doi:10.1136/gut.30.1.78
- Barry J, Gu Y, Jukkola P, O'Neill B, Gu H, Mohler PJ, Rajamani KT, Gu C. 2014. Ankyrin-G directly binds to kinesin-1 to transport voltage-gated Na⁺ channels into axons. *Dev Cell* **28**:117–31. doi:10.1016/j.devcel.2013.11.023
- Behrendt M, Krahn MP, Al-Bayati H, Amiri M, Rizk S, Naim HY. 2012. Cadherin-related protein 24 induces morphological changes and partial cell polarization by facilitating direct cell-cell interactions. *Biol Chem* **393**. doi:10.1515/hsz-2011-0203
- Bitner-Glindzicz M, Glaser B, Lindley KJ, Rutland P, Blaydon D, Smith V V., Milla PJ, Hussain K, Furth-Lavi J, Cosgrove KE, Shepherd RM, Barnes PD, O'Brien RE, Farndon PA, Sowden J, Liu X-Z, Scanlan MJ, Malcolm S, Dunne MJ, Aynsley-Green A. 2000. A recessive contiguous gene deletion causing infantile hyperinsulinism, enteropathy and deafness identifies the Usher type 1C gene. *Nat Genet* **26**:56–60. doi:10.1038/79178
- Bork P, Patthy L. 1995. The SEA module: A new extracellular domain associated with O -glycosylation. *Protein Sci*

4:1421–1425. doi:10.1002/pro.5560040716

- Brasch J, Goodman KM, Noble AJ, Rapp M, Mannepalli S, Bahna F, Dandey VP, Bepler T, Berger B, Maniatis T, Potter CS, Carragher B, Honig B, Shapiro L. 2019. Visualization of clustered protocadherin neuronal self-recognition complexes. *Nature* **1**. doi:10.1038/s41586-019-1089-3
- Brasch J, Harrison OJ, Ahlsen G, Carnally SM, Henderson RM, Honig B, Shapiro L. 2011. Structure and binding mechanism of vascular endothelial cadherin: a divergent classical cadherin. *J Mol Biol* **408**:57–73. doi:10.1016/j.jmb.2011.01.031
- Brasch J, Harrison OJ, Honig B, Shapiro L. 2012. Thinking outside the cell: how cadherins drive adhesion. *Trends Cell Biol* **22**:299–310. doi:10.1016/j.tcb.2012.03.004
- Bujko M, Kober P, Statkiewicz M, Mikula M, Ligaj M, Zwierzchowski L, Ostrowski J, Siedlecki Janusz A. 2015. Epigenetic-Mediated Downregulation of μ -Protocadherin in Colorectal Tumours. *Gastroenterol Res Pract* **2015**:1–9. doi:10.1155/2015/317093
- Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N. 2013. ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Isr J Chem* **53**:199–206. doi:10.1002/ijch.201200096
- Choudhary D, Narui Y, Neel BL, Wimalasena LN, Klanseck CF, De-la-Torre P, Chen C, Araya-Secchi R, Tamilselvan E, Sotomayor M. 2019. Structural determinants of protocadherin-15 elasticity and function in inner-ear mechanotransduction. *bioRxiv* 695502. doi:10.1101/695502
- Chuang S-C, Hsi E, Wang S-N, Yu M-L, Lee K-T, Juo S-HH. 2011. Polymorphism at the mucin-like protocadherin gene influences susceptibility to gallstone disease. *Clin Chim Acta* **412**:2089–2093. doi:10.1016/j.cca.2011.07.015
- Cooper SR, Jontes JD, Sotomayor M. 2016. Structural determinants of adhesion by Protocadherin-19 and implications for its role in epilepsy. *Elife* **5**. doi:10.7554/eLife.18529
- Cowtan K. 2007. Fitting molecular fragments into electron density. *Acta Crystallogr Sect D Biol Crystallogr* **64**:83–89. doi:10.1107/S0907444907033938
- Cowtan K. 2006. The *Buccaneer* software for automated model building. 1. Tracing protein chains. *Acta Crystallogr Sect D Biol Crystallogr* **62**:1002–1011. doi:10.1107/S0907444906022116
- Crawley SW, Mooseker MS, Tyska MJ. 2014a. Shaping the intestinal brush border. *J Cell Biol* **207**:441–51.

doi:10.1083/jcb.201407015

- Crawley SW, Shifrin DA, Grega-Larson NE, McConnell RE, Benesh AE, Mao S, Zheng Y, Zheng QY, Nam KT, Millis BA, Kachar B, Tyska MJ. 2014b. Intestinal brush border assembly driven by protocadherin-based intermicrovillar adhesion. *Cell* **157**:433–446. doi:10.1016/j.cell.2014.01.067
- Crawley SW, Weck ML, Grega-Larson NE, Shifrin DA, Tyska MJ. 2016. ANKS4B Is Essential for Intermicrovillar Adhesion Complex Formation. *Dev Cell* **36**:190–200. doi:10.1016/j.devcel.2015.12.022
- De-la-Torre P, Choudhary D, Araya-Secchi R, Narui Y, Sotomayor M. 2018. A Mechanically Weak Extracellular Membrane-Adjacent Domain Induces Dimerization of Protocadherin-15. *Biophys J* **115**:2368–2385. doi:10.1016/J.BPJ.2018.11.010
- Dickinson DJ, Weis WI, Nelson WJ. 2011. Protein Evolution in Cell and Tissue Development: Going Beyond Sequence and Transcriptional Analysis. *Dev Cell* **21**:32–34. doi:10.1016/j.devcel.2011.06.004
- Dionne G, Qiu X, Rapp M, Liang X, Zhao B, Peng G, Katsamba PS, Ahlsen G, Rubinstein R, Potter CS, Carragher B, Honig B, Müller U, Shapiro L. 2018. Mechanotransduction by PCDH15 Relies on a Novel cis-Dimeric Architecture. *Neuron* **99**:480–492.e5. doi:10.1016/j.neuron.2018.07.006
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792–1797. doi:10.1093/nar/gkh340
- Elledge HM, Kazmierczak P, Clark P, Joseph JS, Kolatkar A, Kuhn P, Müller U. 2010. Structure of the N terminus of cadherin 23 reveals a new adhesion mechanism for a subset of cadherin superfamily members. *Proc Natl Acad Sci U S A* **107**:10708–10712. doi:10.1073/pnas.1006284107
- Emond MR, Biswas S, Blevins CJ, Jontes JD. 2011. A complex of Protocadherin-19 and N-cadherin mediates a novel mechanism of cell adhesion. *J Cell Biol* **195**:1115–21. doi:10.1083/jcb.201108115
- Emond MR, Jontes JD. 2014. Bead aggregation assays for the characterization of putative cell adhesion molecules. *J Vis Exp* e51762. doi:10.3791/51762
- Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**:486–501. doi:10.1107/S0907444910007493
- Frank M, Kemler R. 2002. Protocadherins. *Curr Opin Cell Biol* **14**:557–62.
- Franz F, Daday C, Gräter F. 2020. Advances in molecular simulations of protein mechanical properties and function. *Curr*

1 *Opin Struct Biol* **61**:132–138. doi:10.1016/j.sbi.2019.12.015

2 Ge J, Elferich J, Goehring A, Zhao J, Schuck P, Gouaux E. 2018. Structure of mouse protocadherin 15 of the stereocilia

3 tip link in complex with LHFPL5. *Elife* **7**. doi:10.7554/eLife.38770

4 Gilson AI, Marshall-Christensen A, Choi J-M, Shakhnovich EI. 2017. The Role of Evolutionary Selection in the

5 Dynamics of Protein Structure Evolution. *Biophys J* **112**:1350–1365. doi:10.1016/J.BPJ.2017.02.029

6 Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. 2003. ConSurf: identification of functional

7 regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**:163–4.

8 doi:10.1093/bioinformatics/19.1.163

9 Goldberg M, Peshkovsky C, Shifteh A, Al-Awqati Q. 2000. μ -Protocadherin, a Novel Developmentally Regulated

10 Protocadherin with Mucin-like Domains. *J Biol Chem* **275**:24622–24629. doi:10.1074/jbc.M000234200

11 Goldberg M, Wei M, Tycko B, Falikovich I, Warburton D. 2002. Identification and expression analysis of the human μ -

12 protocadherin gene in fetal and adult kidneys. *Am J Physiol Physiol* **283**:F454–F463.

13 doi:10.1152/ajprenal.00012.2002

14 Goodman KM, Rubinstein R, Thu CA, Bahna F, Mannepalli S, Ahlsén G, Rittenhouse C, Maniatis T, Honig B, Shapiro L.

15 2016. Structural Basis of Diverse Homophilic Recognition by Clustered α - and β -Protocadherins. *Neuron*.

16 doi:10.1016/j.neuron.2016.04.004

17 Grubmüller H. 2005. Force probe molecular dynamics simulations. *Methods Mol Biol* **305**:493–515.

18 Gumbiner BM. 2005. Regulation of cadherin-mediated adhesion in morphogenesis. *Nat Rev Mol Cell Biol* **6**:622–34.

19 doi:10.1038/nrm1699

20 Harrison OJ, Brasch J, Katsamba PS, Ahlsen G, Noble AJ, Dan H, Sampogna R V., Potter CS, Carragher B, Honig B,

21 Shapiro L. 2020. Family-wide Structural and Biophysical Analysis of Binding Interactions among Non-clustered δ -

22 Protocadherins. *Cell Rep* **30**:2655–2671.e7. doi:10.1016/j.celrep.2020.02.003

23 Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A*

24 **89**:10915–9. doi:10.1073/pnas.89.22.10915

25 Huang J, MacKerell ADJ. 2014. CHARMM36 all-atom additive protein force field: Validation based on comparison to

26 NMR data. *J Comput Chem* **34**:2135–2145. doi:10.1002/jcc.23354.CHARMM36

27 Hulpiau P, van Roy F. 2009. Molecular evolution of the cadherin superfamily. *Int J Biochem Cell Biol* **41**:349–369.

doi:10.1016/j.biocel.2008.09.027

Humphrey W, Dalke A, Schulten K. 1996. VMD: Visual molecular dynamics. *J Mol Graph* **14**:33–38. doi:10.1016/0263-7855(96)00018-5

Hussain K, Bitner-Glindzicz M, Blaydon D, Lindley KJ, Thompson DA, Kriss T, Rajput K, Ramadan DG, Al-Mazidi Z, Cosgrove KE, Dunne MJ, Aynsley-Green A. 2004. Infantile hyperinsulinism associated with enteropathy, deafness and renal tubulopathy: clinical manifestations of a syndrome caused by a contiguous gene deletion located on chromosome 11p. *J Pediatr Endocrinol Metab* **17**:1613–21. doi:10.1515/jpem.2004.17.12.1613

Isralewitz B, Gao M, Schulten K. 2001. Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol* **11**:224–230.

Jaiganesh A, De-la-Torre P, Patel AA, Termine DJ, Velez-Cortes F, Chen C, Sotomayor M. 2018. Zooming in on Cadherin-23: Structural Diversity and Potential Mechanisms of Inherited Deafness. *Structure*. doi:10.1016/j.str.2018.06.003

Jiang M, Chen G. 2006. High Ca²⁺-phosphate transfection efficiency in low-density neuronal cultures. *Nat Protoc* **1**:695–700. doi:10.1038/nprot.2006.86

Jin X, Walker M a, Felsövályi K, Vendome J, Bahna F, Mannepalli S, Cosmanescu F, Ahlsen G, Honig B, Shapiro L. 2012. Crystal structures of Drosophila N-cadherin ectodomain regions reveal a widely used class of Ca²⁺-free interdomain linkers. *Proc Natl Acad Sci U S A* **109**:E127–34. doi:10.1073/pnas.1117538108

Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. 2012. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* **7**:1511–1522. doi:10.1038/nprot.2012.085

Kazmierczak P, Sakaguchi H, Tokita J, Wilson-Kubalek EM, Milligan RA, Müller U, Kachar B. 2007. Cadherin 23 and protocadherin 15 interact to form tip-link filaments in sensory hair cells. *Nature* **449**:87–91. doi:10.1038/nature06091

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**:1647–9. doi:10.1093/bioinformatics/bts199

Khubchandani SR, Vohra P, Chitale AR, Sidana P. 2011. Microvillous inclusion disease--an ultrastructural diagnosis:

with a review of the literature. *Ultrastruct Pathol* **35**:87–91. doi:10.3109/01913123.2010.537438

Krahn MP, Rizk S, Alfalah M, Behrendt M, Naim HY. 2010. Protocadherin of the Liver, Kidney, and Colon Associates with Detergent-resistant Membranes during Cellular Differentiation. *J Biol Chem* **285**:13193–13200. doi:10.1074/jbc.M109.080051

Kwon M, Firestein BL. 2013. DNA transfection: calcium phosphate method. *Methods Mol Biol* **1018**:107–10. doi:10.1007/978-1-62703-444-9_10

Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. 2005. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* **33**:W299-302. doi:10.1093/nar/gki370

Lefebvre JL. 2017. Neuronal territory formation by the atypical cadherins and clustered protocadherins. *Semin Cell Dev Biol* **69**. doi:10.1016/J.SEMCDB.2017.07.040

Li J, He Y, Lu Q, Zhang M. 2016. Mechanistic Basis of Organization of the Harmonin/USH1C-Mediated Brush Border Microvilli Tip-Link Complex. *Dev Cell* **36**:179–89. doi:10.1016/j.devcel.2015.12.020

Losi L, Grande A. 2014. Involvement of μ -protocadherin in colorectal carcinogenesis: a promise for clinico-pathological evaluation. *Histol Histopathol* **29**:11–9. doi:10.14670/HH-29.11

Maroux S, Coudrier E, Feracci H, Gorvel J-P, Louvard D. 1988. Molecular organization of the intestinal brush border. *Biochimie* **70**:1297–1306. doi:10.1016/0300-9084(88)90198-8

McConnell RE, Benesh AE, Mao S, Tabb DL, Tyska MJ. 2011. Proteomic analysis of the enterocyte brush border. *Am J Physiol Gastrointest Liver Physiol* **300**:G914-26. doi:10.1152/ajpgi.00005.2011

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* **40**:658–674. doi:10.1107/S0021889807021206

Modak D, Sotomayor M. 2019. Identification of an adhesive interface for the non-clustered $\delta 1$ protocadherin-1 involved in respiratory diseases. *Commun Biol* **2**:354. doi:10.1038/s42003-019-0586-0

Mooseker MS. 1985. Organization, chemistry, and assembly of the cytoskeletal apparatus of the intestinal brush border. *Annu Rev Cell Biol* **1**:209–41. doi:10.1146/annurev.cb.01.110185.001233

Mukherjee S, Vaishnava S, Hooper L V. 2008. Multi-layered regulation of intestinal antimicrobial defense. *Cell Mol Life Sci* **65**:3019–3027. doi:10.1007/s00018-008-8182-3

- 1 Murshudov GN, Skubák P, Lebedev A a., Pannu NS, Steiner R a., Nicholls R a., Winn MD, Long F, Vagin A a. 2011.
- 2 REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr Sect D Biol Crystallogr*
- 3 **67**:355–367. doi:10.1107/S0907444911001314
- 4 Nagar B, Overduin M, Ikura M, Rini JM. 1996. Structural basis of calcium-induced E-cadherin rigidification and
- 5 dimerization. *Nature* **380**:360–4. doi:10.1038/380360a0
- 6 Nicoludis JM, Green AG, Walujkar S, May EJ, Sotomayor M, Marks DS, Gaudet R. 2019. Interaction specificity of
- 7 clustered protocadherins inferred from sequence covariation and structural analysis. *Proc Natl Acad Sci U S A*
- 8 **116**:17825–17830. doi:10.1073/pnas.1821063116
- 9 Nicoludis JM, Lau S-Y, Schärfe CPI, Marks DS, Weihofen WA, Gaudet R. 2015. Structure and Sequence Analyses of
- 10 Clustered Protocadherins Reveal Antiparallel Interactions that Mediate Homophilic Specificity. *Structure* **23**:2087–
- 11 98. doi:10.1016/j.str.2015.09.005
- 12 Nicoludis JM, Vogt BE, Green AG, Schärfe CP, Marks DS, Gaudet R. 2016. Antiparallel protocadherin homodimers use
- 13 distinct affinity- and specificity-mediating regions in cadherin repeats 1-4. *Elife* **5**. doi:10.7554/eLife.18449
- 14 Okazaki N, Takahashi N, Kojima S-I, Masuho Y, Koga H. 2002. Protocadherin LKC, a new candidate for a tumor
- 15 suppressor of colon and liver cancers, its association with contact inhibition of cell proliferation. *Carcinogenesis*
- 16 **23**:1139–1148. doi:10.1093/carcin/23.7.1139
- 17 Ose R, Yanagawa T, Ikeda S, Ohara O, Koga H. 2009. PCDH24-induced contact inhibition involves downregulation of β -
- 18 catenin signaling. *Mol Oncol* **3**:54–66. doi:10.1016/j.molonc.2008.10.005
- 19 Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*
- 20 **276**:307–26.
- 21 Parisini E, Higgins JMG, Liu JH, Brenner MB, Wang JH. 2007. The crystal structure of human E-cadherin domains 1 and
- 22 2, and comparison with other cadherins in the context of adhesion mechanism. *J Mol Biol* **373**:401–11.
- 23 doi:10.1016/j.jmb.2007.08.011
- 24 Pei J, Grishin N V. 2017. Expansion of divergent SEA domains in cell surface proteins and nucleoporin 54. *Protein Sci*
- 25 **26**:617–630. doi:10.1002/pro.3096
- 26 Pepermans E, Petit C. 2015. The tip-link molecular complex of the auditory mechano-electrical transduction machinery.
- 27 *Hear Res* **330**:10–7. doi:10.1016/j.heares.2015.05.005

- 1 Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. 2005.
- 2 Scalable molecular dynamics with NAMD. *J Comput Chem* **26**:1781–1802. doi:10.1002/jcc.20289
- 3 Pinette JA, Mao S, Millis BA, Krystofiak ES, Faust JJ, Tyska MJ. 2018. Brush border protocadherin CDHR2 promotes
- 4 the elongation and maximized packing of microvilli in vivo. *Mol Biol Cell* **30**:108–118. doi:10.1091/mbc.e18-09-
- 5 0558
- 6 Ponstingl H, Henrick K, Thornton JM. 2000. Discriminating between homodimeric and monomeric proteins in the
- 7 crystalline state. *Proteins* **41**:47–57.
- 8 Powers RE, Gaudet R, Sotomayor M. 2017. A Partial Calcium-Free Linker Confers Flexibility to Inner-Ear
- 9 Protocadherin-15. *Structure* **25**:482–495. doi:10.1016/j.str.2017.01.014
- 10 Rubinstein R, Thu CA, Goodman KM, Wolcott HN, Bahna F, Manneppalli S, Ahlsen G, Chevee M, Halim A, Clausen H,
- 11 Maniatis T, Shapiro L, Honig B. 2015. Molecular Logic of Neuronal Self-Recognition through Protocadherin
- 12 Domain Interactions. *Cell* **163**:629–42. doi:10.1016/j.cell.2015.09.026
- 13 Sano K, Tanihara H, Heimark RL, Obata S, Davidson M, St John T, Taketani S, Suzuki S. 1993. Protocadherins: a large
- 14 family of cadherin-related molecules in central nervous system. *EMBO J* **12**:2249–56.
- 15 Schreiner D, Weiner JA. 2010. Combinatorial homophilic interaction between gamma-protocadherin multimers greatly
- 16 expands the molecular diversity of cell adhesion. *Proc Natl Acad Sci U S A* **107**:14893–8.
- 17 doi:10.1073/pnas.1004526107
- 18 Selsted ME, Ouellette AJ. 2005. Mammalian defensins in the antimicrobial immune response. *Nat Immunol* **6**:551–557.
- 19 doi:10.1038/ni1206
- 20 Shapiro L, Fannon AM, Kwong PD, Thompson A, Lehmann MS, Grubel G, Legrand JF, Als-Nielsen J, Colman DR,
- 21 Hendrickson WA. 1995. Structural basis of cell-cell adhesion by cadherins. *Nature* **374**:327–37.
- 22 doi:10.1038/374327a0
- 23 Shifrin DA, Tyska MJ. 2012. Ready...aim...fire into the lumen: a new role for enterocyte microvilli in gut host defense.
- 24 *Gut Microbes* **3**:460–2. doi:10.4161/gmic.21247
- 25 Sotomayor M, Gaudet R, Corey DP. 2014. Sorting out a promiscuous superfamily: towards cadherin connectomics.
- 26 *Trends Cell Biol* **24**:524–36. doi:10.1016/j.tcb.2014.03.007
- 27 Sotomayor M, Schulten K. 2007. Single-molecule experiments in vitro and in silico. *Science* **316**:1144–1148.

doi:10.1126/science.1137591

- Sotomayor M, Weihofen WA, Gaudet R, Corey DP. 2012. Structure of a force-conveying cadherin bond essential for inner-ear mechanotransduction. *Nature* **492**:128–32. doi:10.1038/nature11590
- Sotomayor M, Weihofen WA, Gaudet R, Corey DP. 2010. Structural Determinants of Cadherin-23 Function in Hearing and Deafness. *Neuron* **66**:85–100. doi:10.1016/j.neuron.2010.03.028
- Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT-BG, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, Brunak S, Wandall HH, Levery SB, Clausen H. 2013. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* **32**:1478–88. doi:10.1038/emboj.2013.79
- Takeichi M. 1990. Cadherins: a molecular family important in selective cell-cell adhesion. *Annu Rev Biochem* **59**:237–52. doi:10.1146/annurev.bi.59.070190.001321
- Vallance BA, Chan C, Robertson ML, Finlay BB. 2002. Enteropathogenic and enterohemorrhagic Escherichia coli infections: emerging themes in pathogenesis and prevention. *Can J Gastroenterol* **16**:771–8. doi:10.1155/2002/410980
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**:1189–1191. doi:10.1093/bioinformatics/btp033
- Weck ML, Crawley SW, Stone CR, Tyska MJ. 2016. Myosin-7b Promotes Distal Tip Localization of the Intermicrovillar Adhesion Complex. *Curr Biol* **26**:2717–2728. doi:10.1016/j.cub.2016.08.014
- Wilson W, Scott RB, Pinto A, Robertson MA. 2001. Intractable diarrhea in a newborn infant: microvillous inclusion disease. *Can J Gastroenterol* **15**:61–4. doi:10.1155/2001/743925

Figure 1. Comparison of sequence conservation across species. (A) Percent identity across the extracellular domains of PCDH24 (EC1-MAD10), CDH23 (EC1-MAD28), CDHR5 (EC1-4), and PCDH15 (EC1-MAD12) plotted against EC repeat numbers. Overall the cadherins of the inner-ear tip link, CDH23 and PCDH15, have higher identity across a variety of species while the intermicrovillar-link cadherins, PCDH24 and CDHR5, have poor sequence conservation across species (Figure S1, S2; Tables S2-S5). N- and C-termini are more conserved in CDH23 and PCDH15 than the middle region of these proteins. An opposite trend is observed for PCDH24 and CDHR5. (B) Percent identity of the first three EC repeats of PCDH24, CDHR5, CDH23, PCDH15, CDH1 and CDH2 demonstrate that PCDH24 and CDHR5 sequences are not highly conserved across several species (*Homo sapiens*, *hs*; *Mus musculus*, *mm*; *Gallus gallus*, *gg*; *Anolis carolinensis*, *ac*; *Danio rerio*, *dr*). However, the N-terminal repeats of CDH23, PCDH15, CDH1, and CDH2 are highly conserved. Sequences were obtained from NCBI (Table S6 and Methods), except for *gg* CDHR5, which is unavailable.

Figure 2. Structures of human (*hs*) PCDH24 EC1-2, mouse (*mm*) PCDH24 EC1-3, and human (*hs*) CDHR5 EC1-2. (A) Topology of *hs* PCDH24 EC1-2. The EC1 and EC2 repeats have a typical cadherin fold with 7 β -strands. A disulfide bond in EC2 is highlighted in orange. Calcium ions are shown as green circles. (B) A ribbon representation of *hs* PCDH24 EC1-2 with calcium ions in green. (C) Topology of the *mm* PCDH24 EC1-3 shown as in (A). The dashed line indicates a loop not resolved. (D) A ribbon representation of *mm* PCDH24 EC1-3 as in (B). The structure is similar to *hs* PCDH24 EC1-2, but the N-terminus projects away from EC1. A N-terminal calcium ion was not present. (E) Detail of the linker between EC1 and EC2 of *hs* PCDH24 showing canonical calcium-binding sites. Some sidechain and backbone atoms are not shown for clarity. (F) Detail of the *mm* PCDH24 EC2-3 linker, which lacks several canonical calcium-binding residues and thus only has two calcium ions bound. Shown as in (E). An asterisk (*) indicates the position where a calcium residue would be found in the canonical linker. (G) Detail of the N-terminal calcium bound at the tip of EC1 in the *hs* PCDH24 EC1-2 structure. A similar calcium-binding motif is seen at the tip of CDH23. Residues coordinating calcium are shown in stick with backbone atoms omitted. (H) Detail of the N-terminus of *mm* PCDH24 EC1-3 (blue) showing its interaction with another protomer (cyan) in the asymmetric unit. (I) Topology of *hs* CDHR5 EC1-2 shown as in (A). (J) A ribbon representation of *hs* CDHR5 EC1-2 with calcium ions in green. (K) Detail of the linker between EC1 and EC2 of *hs* CDHR5 showing canonical calcium-binding sites. Shown as in (I-J). (L) Detail of the disulfide bond in EC1 in the *hs* CDHR5 EC1-2 structure. A similar disulfide bond is seen in PCDH15 EC1.

Figure 3. Homophilic binding assays of *hs* PCDH24. (A-H) Protein G beads coated with the full-length *hs* PCDH24 extracellular domain (A) and its C-terminal truncation versions (B-H). Each image shows bead aggregation observed after 60 min in the presence of 2 mM CaCl_2 . (I-J) Protein G beads coated with the full-length *hs* PCDH24 extracellular domain (I) and EC1 (J) in the presence of 2 mM EDTA, shown as in (A-H). (K) Detail of binding assay results for EC1-2Fc, EC1Fc, EC1-MAD10Fc with EDTA, and EC1Fc with EDTA. Aggregation is still present for the shortest constructs. Bar – 500 μm . (L) Aggregate size for full-length and truncated versions of *hs* PCDH24 extracellular domains at the start of the experiment (T0), and after 30 and 60 min (T30, T60). Inset shows aggregate size of the shortest constructs compared to the EDTA control of EC1-MAD10Fc. An additional one-minute rocking step is denoted by R1. Error bars are standard error of the mean (n indicated in Table S7). (M) Western blot shows expression and secretion of full-length and truncated versions of *hs* PCDH24.

Figure 4. Comparison of homophilic binding assays of *hs* and *mm* PCDH24 and CDHR5. (A-D) Protein G beads coated with full-length *hs* PCDH24 (A), full-length *mm* PCDH24 (B), full-length *hs* CDHR5 (C), and full-length *mm* CDHR5 (D) cadherin extracellular domains, including MAD10 for PCDH24. None of the CDHR5 fragments included the MLD. Images show the aggregation observed at the start of the experiment (T0), after 60 min (T60) followed by rocking for 1 min (R1) in the presence of 2 mM CaCl₂. Bar = 500 μm. (E) Aggregate size for full-length *hs* and *mm* PCDH24 extracellular domains in the presence of 2 mM CaCl₂ and 2 mM EDTA at the start of the experiment (T0), after 60 min (T60) followed by rocking for 1 minute (R1). Inset shows aggregate size of full-length *mm* PCDH24 extracellular domain compared to the EDTA control of full-length *hs* and *mm* PCDH24 extracellular domains. (F) Aggregate size for full-length *hs* and *mm* CDHR5 cadherin extracellular domains in the presence of 2 mM CaCl₂ and 2 mM EDTA at the start of the experiment (T0), after 60 min (T60) followed by rocking for 1 min (R1). Inset shows aggregate size of the full-length *hs* CDHR5 cadherin extracellular domain compared to the EDTA control of full-length *hs* and *mm* CDHR5 cadherin extracellular domain. Error bars in (E) and (F) are standard error of the mean (*n* indicated in Table S7). (G) Western blot shows expression and secretion of full-length *hs* and *mm* PCDH24 and CDHR5 cadherin extracellular domains.

Figure 5. Homophilic binding assays of *mm* CDHR5. (A-F) Protein G beads coated with the full-length *mm* CDHR5 cadherin extracellular domain (A) and its C-terminal truncation versions (B-D). Each image shows bead aggregation observed after 60 min followed by rocking for 2 min in the presence of 2 mM CaCl₂. (E-F) Protein G beads coated with mutants *mm* CDHR5 EC1-4Fc R82G (E) and EC1-4Fc E84G (F) observed after 60 min followed by rocking for 2 min in the presence of 2 mM CaCl₂. (G) Protein G beads coated with the full-length *mm* CDHR5 cadherin extracellular domain in the presence of 2 mM EDTA shown as in (A). Bar – 500 μm. (H) Aggregate size for full-length, truncated versions and mutants of *mm* CDHR5 cadherin extracellular domains at the start of the experiment (T0), after 60 min (T60) followed by rocking for 1 min (R1) and 2 min (R2). Inset shows aggregate size of EC1Fc, EC1-2Fc, the EDTA control of EC1-4Fc and EC1-4Fc mutants R82G and E84G. Error bars are standard error of the mean (*n* indicated in Table S7). (I) Western blot shows expression and secretion of full-length, truncated, and mutant versions of *mm* CDHR5 cadherin extracellular domains.

1

Figure 6. Heterophilic binding assays of *hs* and *mm* PCDH24 and CDHR5. (A-B) Images from binding assays of *hs* PCDH24 (full-length extracellular domain and EC1-4Fc) mixed with the full-length *hs* CDHR5 cadherin extracellular domain (A), and *mm* PCDH24 (full-length extracellular domain and EC1-4Fc) mixed with full-length *mm* CDHR5 cadherin extracellular domain (B). Green fluorescent protein A beads are coated with PCDH24 fragments and red fluorescent protein A beads are coated with CDHR5 in all panels. All images are details of complete views shown in Figure S17. Images show bead aggregation observed after 60 min followed by rocking for 2 min in the presence of 2 mM CaCl₂. (C-F) Images from binding assays of truncations of *hs* CDHR5 mixed with *hs* PCDH24 EC1-4Fc (C), truncations of *hs* PCDH24 mixed with *hs* CDHR5 EC1-4Fc (D), truncations of *mm* CDHR5 mixed with *mm* PCDH24 EC1-4Fc (E), and truncations of *mm* PCDH24 mixed with *mm* CDHR5 EC1-4Fc (F). Images show bead aggregation observed after 60 min followed by rocking for 2 min in the presence of 2 mM CaCl₂. (G) Images from binding assays of minimum EC repeats required for heterophilic adhesion of *hs* and *mm* PCDH24 and CDHR5. The minimum units for heterophilic adhesion for the human proteins are CDHR5 EC1Fc and PCDH24 EC1-2Fc. The minimum units for heterophilic adhesion for the mouse proteins are CDHR5 EC1-2Fc and PCDH24 EC1-2Fc. Images show bead aggregation observed after 60 minutes followed by rocking for 2 minutes in the presence of 2 mM CaCl₂. (H) Protein A beads coated with full-length *hs* and *mm* PCDH24 and CDHR5 cadherin extracellular domains (including MAD10 for PCDH24 and without the MLD for CDHR5) in the presence of 2 mM EDTA shown as in (A-B). Bar – 500 μm. (I) Aggregate size for full-length and truncated constructs of *hs* PCDH24 and *hs* CDHR5 cadherin extracellular domains (including MAD10 for PCDH24 when indicated and without the mucin-like domain for CDHR5) at the start of the experiment (T0), after 60 min (T60) followed by rocking for 1 min (R1) and 2 min (R2). (J) Aggregate size for full-length and truncated constructs of *mm* PCDH24 and *mm* CDHR5 as in (I). Error bars in I and J are standard error of the mean (*n* indicated in

- 1 Table S7). (K) Western blot shows expression and secretion of full-length and truncated versions of *hs* and *mm* PCDH24 and CDHR5
- 2 cadherin extracellular domains (including MAD10 for PCDH24 when indicated and without the MLD for CDHR5).

Figure 7. Schematics of homophilic and heterophilic adhesion mediated by human and mouse PCDH24 and CDHR5. Homophilic adhesion that results in small bead aggregates can be mediated by *hs* PCDH24 EC1 (calcium independent) and EC1-2 (calcium dependent), while homophilic adhesion that results in large bead aggregates requires *hs* PCDH24 EC1-3 (calcium dependent). Homophilic adhesion mediated by *mm* CDHR5 requires EC1-3 (calcium dependent). Heterophilic adhesion is mediated by PCDH24 EC1-2 and CDHR5 EC1 when using human proteins and by PCDH24 EC1-2 and CDHR5 EC1-2 when using mouse proteins (both calcium dependent).

1

Table 1. Statistics for PCDH24 and CDHR5 structures			
Data Collection and Refinement	<i>Hs</i> PCDH24 EC1-2	<i>Mm</i> PCDH24 EC1-3	<i>Hs</i> CDHR5 EC1-2
Space Group	P2 ₁	P321	I222
Unit cell parameters			
a, b, c (Å)	270.9, 119.8, 74.3	74.8, 74.8, 140.0	50.4, 78.4, 121.4
α, β, γ (°)	90, 104.12, 90	90, 90, 120	90, 90, 90
Molecules per asymmetric unit	4	1	1
Beam Source	APS 24-ID-C	APS 24-ID-C	APS 24-ID-E
Wavelength (Å)	0.97910	0.97910	0.97918
Resolution limit (Å)	2.3	2.1	1.9
Unique Reflections	50945 (2128)	27040 (1237)	19398 (943)
Redundancy	3.1 (2.6)	8.60 (5.20)	24.10 (21.80)
Completeness (%)	95.5 (80.3)	99.2 (90.6)	99.7 (99.7)
Average I/σ (I)	10.9 (3)	27 (4.4)	46.14 (14.50)
R _{merge}	0.105 (0.255)	0.078 (0.337)	0.074 (0.263)
Refinement			
Resolution range (Å)	50 – 2.3 (2.34 – 2.3)	50 – 2.1 (2.14 – 2.1)	50-1.9 (1.93-1.9)
Residues (atoms)	861 (6610)	318 (2447)	204 (1656)
Water Molecules	238	98	128
R _{work} (%)	20.5 (27.8)	21.0 (34.3)	19.7 (20.9)
R _{free} (%)	23.5 (33.0)	24.3 (39.6)	23.4 (25.9)
Rms deviations			
Bond lengths (Å)	0.013	0.014	0.013
Bond angles (°)	1.460	1.632	1.652
B factor average			
Protein	41.00	64.31	30.53
Ligand/ion	36.90	51.49	39.42
Water	31.65	50.99	33.6
Ramachandran plot regions			
Most favored (%)	91.5	90.7	91.2
Additionally allowed (%)	8.3	8.9	8.8
Generously allowed (%)	0.3	0.4	0.0
Disallowed (%)	0.0	0.0	0.0
PDB ID code	5CZR	5CYX	6OAE

2

SUPPLEMENTARY MATERIAL

SPECIES-DEPENDENT HETEROPHILIC AND HOMOPHILIC CADHERIN INTERACTIONS IN INTESTINAL INTERMICROVILLAR LINKS

Michelle E. Gray*¹, Zachary R. Johnson*^{§2}, Debadrita Modak*², Matthew J. Tyska³, and Marcos Sotomayor*^{#1,2}

¹Ohio State Biochemistry Program

²Department of Chemistry and Biochemistry

The Ohio State University, 484 W 12th Avenue, Columbus, OH 43210

³Department of Cell and Developmental Biology

Vanderbilt University School of Medicine, Nashville, TN 37232

*Co-first authors

[§]Current address: Marshall University, Joan C. Edwards School of Medicine, Huntington, WV 25701

Corresponding Author

E-mail: sotomayor.8@osu.edu

Figure S1. Sequence alignments of individual EC repeats of PCDH24. Multiple sequence alignments comparing each EC repeat and MAD10 of PCDH24 from 19 different species. Each alignment is colored by percent identity, with white being the lowest percent identity and dark blue being the highest. Sites of N-linked (blue in human, green in mouse) and O-linked (purple in human, red in mouse) glycosylation are denoted by a colored circle. Secondary structure elements observed in the crystal structures of *hs* PCDH24 EC1-2 and *mm* PCDH24 EC1-3 are illustrated below the respective repeats. Calcium-binding motifs are indicated above the sequences, which are numbered according to the human protein. Species are abbreviated as follows: *Homo sapiens* (*hs*), *Pan troglodytes* (*Pt*), *Rattus norvegicus* (*Rn*), *Mus musculus* (*mm*), *Canis lupus familiaris* (*Clf*), *Felis catus* (*Fc*), *Acinonyx jubatus* (*Aj*), *Bos taurus* (*Bt*), *Sus scrofa* (*Ss*), *Ovis aries* (*Oa*), *Phascolarctos cinereus* (*Pc*), *Delphinapterus leucas* (*Dl*), *Gallus gallus* (*Gg*), *Corvus cornix* (*Ccc*), *Aptenodytes forsteri* (*Af*), *Anolis carolinensis* (*Ac*), *Danio rerio* (*Dr*), *Astyanax mexicanus* (*Am*), and *Oncorhynchus mykiss* (*Om*). Species were chosen based on sequence availability and taxonomical diversity. Accession numbers and species can be found in Table S2.

Figure S2. Sequence alignments of individual EC repeats of CDHR5. Multiple sequence alignments comparing each EC repeat of CDHR5 from 15 different species, shown as in Figure S1. An asterisk (*) indicates site R84 mutated in binding assays. An arrow indicates the end of EC1-EC4 protein fragments used in binding assays. Secondary structure elements observed in the crystal structures of *hs* CDHR5 EC1-2 are illustrated below the respective repeats. Calcium-binding motifs are indicated above the sequences, which are numbered according to the human protein. Species are abbreviated as follows: *Homo sapiens* (hs), *Pan troglodytes* (Pt), *Rattus norvegicus* (Rn), *Mus musculus* (mm), *Felis catus* (Fc), *Bos Taurus* (Bt), *Sus scrofa* (Ss), *Ovis aries* (Oa), *Phascolarctos cinereus* (Pc), *Delphinapterus leucas* (DI), *Aptenodytes forsteri* (Af), *Anolis carolinensis* (Ac), *Danio rerio* (Dr), *Astyanax mexicanus* (Am), and *Oncorhynchus mykiss* (Om). Species were chosen based on sequence availability and taxonomical diversity. Accession numbers and species can be found in Table S3.

Figure S3. Comparison of PCDH24 and CDHR5 N-termini with other cadherins. (A) Alignment of the processed N-terminal sequences of Cr-2 protocadherins *hs* PCDH24, *mm* PCDH24, *hs* CDH23 and *hs* PCDH21, and of two classical cadherins, *hs* CDH1 and *hs* CDH2. Calcium-binding sites are labeled and boxed in red in the sequence to show the additional calcium-binding sites in Cr-2 protocadherins not present in classical cadherins. Residue W2 involved in classical cadherin binding is boxed in pink. (B) Alignments of the processed sequences of EC1 for *hs* CDHR5, *mm* CDHR5, and *hs* PCDH15. Calcium-binding sites are labeled. The two cysteine residues that form a disulfide bond between β -strands A and F are highlighted by a red box.

Figure S4. Electron density for unique elements of *hs* PCDH24 EC1-2, *mm* PCDH24 EC1-3 and *hs* CDHR5 EC1-2. (A) Extended F-G loop of EC2 in *hs* PCDH24 where the disulfide bond is located (C187 - C201) shown at 1.0 σ using curve of 1.6. This loop is also present in the *mm* PCDH24 structure. (B) The atypical extended N-terminus seen in *mm* PCDH24 EC1 shown at 1.0 σ and curve of 1.8. (C) Disulfide bond between residue C5 in β -strand A and residue C73 in β -strand F of *hs* CDHR5 EC1-2 shown at 1.0 σ and curve of 1.6.

PCDH24

```

Hs EC1  NVAFKFF-----LANNITSVILPEQLPVGAQAF--WLVAEEQD-----ND-PLTYGM-----SGPNAY-----FAVTPK-----TGEVKLAS--ALDYE---T-LYTFKVTIISVSDP-----YIQVGREMLVIVEDRND---
Mm EC1  NSPSPSF-----GVNMTLVTLPEDLPVGAQAF--WLVAATSD-----ND-HLTYGI-----SGPNAS-----YSVMAN-----TGEVKLAS--ALDYE---T-LYTFKVTIISVSDP-----LNIETAEQVIVEDRND---
Hs EC2  --APVF-----GVNMTLVTLPEDLPVGAQAF--WLVAATSD-----ND-HLTYGI-----SGPNAS-----YSVMAN-----TGEVKLAS--ALDYE---T-LYTFKVTIISVSDP-----LNIETAEQVIVEDRND---
Mm EC2  --APVF-----GVNMTLVTLPEDLPVGAQAF--WLVAATSD-----ND-HLTYGI-----SGPNAS-----YSVMAN-----TGEVKLAS--ALDYE---T-LYTFKVTIISVSDP-----LNIETAEQVIVEDRND---
Hs EC3  --DQFE-----VREYSASVAGAAKAGTSLV--TVEAVSQ--KQIND--PVIYSISYS--TRPG-----WDIGAD-----GVIFVNG--SDREQLLEA--DEEVALQVATET--HLNIVGDEAKASMTIIRVDBVNDH---
Mm EC3  --DQFE-----VREYSASVAGAAKAGTSLV--TVEAVSQ--KQIND--PVIYSISYS--TRPG-----WDIGAD-----GVIFVNG--SDREQLLEA--DEEVALQVATET--HLNIVGDEAKASMTIIRVDBVNDH---
Hs EC4  --KEFYNGSLPACTTPEEAQVNFISGVDEKASARIISIDGITWATPDP--GGDNG--FTLSL-----NGDDAE-----ANVSPER--AAGSVSYQVAVPNSGVNDYE---K-ETWVVEVATID-----SVSNYSVATVTHLRINDH---
Mm EC4  --KEFYNGSLPACTTPEEAQVNFISGVDEKASARIISIDGITWATPDP--GGDNG--FTLSL-----NGDDAE-----ANVSPER--AAGSVSYQVAVPNSGVNDYE---K-ETWVVEVATID-----SVSNYSVATVTHLRINDH---
Hs EC5  --RPVF-----QSIVELTVPEDHPTGVLVTD--KIQATPLD--GGDNG--FTLSL-----NGDDAE-----ANVSPER--AAGSVSYQVAVPNSGVNDYE---K-ETWVVEVATID-----SVSNYSVATVTHLRINDH---
Mm EC5  --RPVF-----QSIVELTVPEDHPTGVLVTD--KIQATPLD--GGDNG--FTLSL-----NGDDAE-----ANVSPER--AAGSVSYQVAVPNSGVNDYE---K-ETWVVEVATID-----SVSNYSVATVTHLRINDH---
Hs EC6  --APVV-----RQSYNVFVPEEN--GVVSV--TIOAYDDQPDPTNNSRLFLNLL--PGPYSH--NLSLDPD-----TGLLRNLG--PLDREAIIDPALLEGRIIVLTVASDC-----GEPVLSTKVNVTIIVEDJNDN---
Mm EC6  --APVV-----RQSYNVFVPEEN--GVVSV--TIOAYDDQPDPTNNSRLFLNLL--PGPYSH--NLSLDPD-----TGLLRNLG--PLDREAIIDPALLEGRIIVLTVASDC-----GEPVLSTKVNVTIIVEDJNDN---
Hs EC7  --LPVF-----NQSYNVFTKEDP--GVLVG--VVKAWMAQTAANN--RISFSL-----SGTGAN-----NIMRGLVLGAGWAEGLYRLPPDVSIDYE---T-QKFFNLTVSAENP-----DPQGETIVDVCVWVKDVND---
Mm EC7  --LPVF-----NQSYNVFTKEDP--GVLVG--VVKAWMAQTAANN--RISFSL-----SGTGAN-----NIMRGLVLGAGWAEGLYRLPPDVSIDYE---T-QKFFNLTVSAENP-----DPQGETIVDVCVWVKDVND---
Hs EC8  --PFTL-----DASLRQIRVAEGSHQGVV--RVIAQGVDTAALLR--IELDVILCTK--AGVDGSLCWMG--SVGAN-----GSVYINQSKAIDYE---A-COLVTLVVRACDL--ATDPQFQAYSSNSSLITIEVDND---
Mm EC8  --PFTL-----DASLRQIRVAEGSHQGVV--RVIAQGVDTAALLR--IELDVILCTK--AGVDGSLCWMG--SVGAN-----GSVYINQSKAIDYE---A-COLVTLVVRACDL--ATDPQFQAYSSNSSLITIEVDND---
Hs EC9  --AVF-----LNNSTFVIIF--VLPNQVVA--SVQARNE-----SEQNG--IIMPSILKAEFYRKDQTSNPVQ--V-----SITRSY--EAGLFISGIELVT--NLQST---L-QDTYQVTVQAGAD-----PLQPALETGTLTLMFTVDSYRNR
Mm EC9  --AVF-----LNNSTFVIIF--VLPNQVVA--SVQARNE-----SEQNG--IIMPSILKAEFYRKDQTSNPVQ--V-----SITRSY--EAGLFISGIELVT--NLQST---L-QDTYQVTVQAGAD-----PLQPALETGTLTLMFTVDSYRNR

```

Figure S5. Sequence alignment of *hs* and *mm* PCDH24 EC repeats. All 9 EC repeats for each species are aligned to each other (EC1-9). Conserved calcium-binding motifs are labeled on top of the alignment.

Figure S6. Equilibrium and stretching simulations of *mm* PCDH24 EC1-3. (A) Superposition of *mm* PCDH24 EC1-3 conformations taken every 5 ns from a ~99 ns long trajectory (simulation S2, Table S10). Repeat EC2 was used as a reference. Side, top, and bottom views are shown. Color indicates time step (red-white-blue). (B) Orientation projections illustrating the conformational freedom of the EC1-2 and EC2-3 linkers throughout equilibrium simulations. To quantify the conformational freedom of EC2 relative to EC1 (top), and of EC3 relative to EC2 (bottom), the longest principal axes of EC1 and EC3 were aligned to the *z* axis and then the projections of the longest principal axes of EC2 (blue) and EC3 (red) in the *x-y* plane were plotted. The initial orientation of a control protomer, CDH23 EC1-2 (2WHV), is shown as a black dot (Sotomayor et al., 2010). The EC2-3 linker behavior is not dramatically different than the behavior observed for the EC1-2 linker, suggesting similar flexibility. (C) Trajectory snapshots during the slowest speed stretching simulation at 0.1 nm/ns for *mm* PCDH24 EC1-3 (S3d, Table S10). Springs indicate position (center of mass) and direction of applied forces. Unfolding of the EC2-3 linker is observed first. End-to-end distances between the centers of mass of EC1 and EC3 are indicated for each snapshot. (D) Force versus end-to-end distance for simulations of the *mm* PCDH24 EC1-3 monomer (S3a-d) at stretching speeds of 10 nm/ns (red), 1 nm/ns (blue) and 0.1 nm/ns (green). Dark and light colors indicate forces applied at opposite ends. Black arrowheads indicate time points illustrated in (C). (E) Detail of the *mm* PCDH24 EC2-3 linker at the force peak during the slowest stretching simulation at 0.1 nm/ns (S3d, Table S10). Residues involved in two-step unbinding from calcium ions are shown.

Figure S7. Structural comparison of PCDH24 and CDHR5 EC1 repeats with other cadherins. (A) Detail of superposed EC1 N-termini of *hs* PCDH24 (blue) and *mm* CDH23 (grey; PDB: 2WHV). Both have a calcium ion at the calcium-binding site 0, a feature of Cr-2 protocadherins. (B) Detail of superposed EC1 N-termini of *mm* PCDH24 (blue) and *hs* CDH1 (grey; PDB: 2O72). Both have N-termini protruding away from the protomer to interact with another neighboring EC1, as shown in (C) for *hs* CDH1 and (D) for *mm* PCDH24. (C-D) Surface representations of *hs* CDH1 and *mm* PCDH24, respectively. Highlighted in yellow is β -strand A with W2 from its binding partner. Residues 25, 26, and 27 are not shown to facilitate visualization of the tryptophan binding pocket. Highlighted in purple is the N-terminus of *mm* PCDH24 from a neighboring protomer in the crystal lattice (Figure 2D). (E) Detail of *hs* CDHR5 EC1 (green) and *mm* PCDH15 EC1 (purple) superposed. Disulfide bonds are shown in orange and yellow. PCDH15 has longer loops and a longer N-terminus, with a helix that interacts with CDH23.

Figure S8. Crystal contacts in the *hs* PCDH24 EC1-2 structure. (A-D) Ribbon diagram of two protomers showing a crystal contact between chains A and D with an interface area of 979.2 Å². The arrangement is antiparallel, likely describing a possible *trans* interface (A). Crystal contacts in the entire asymmetric unit include four additional interfaces with areas of 173.6 Å² (chains A and B), 231.1 Å² (chains A and C), 255.3 Å² (chains D and B), and 307.2 Å² (chains D and C). Two Additional interfaces between chains A and C (387.4 Å² and 507.3 Å²) are shown in (C) and (D). Equivalent interfaces for the one shown in (A) between chains B and C, and those shown in (C, D) between chains D and B, have similar interface areas of 938.6 Å², 390.8 Å², and 515.1 Å², respectively, but are not shown.

Figure S9. Stretching simulations testing the strength of the *hs* PCDH24 EC1-2 *trans* interface. (A) Force versus end-to-end distance for simulations of the largest crystallographic *trans* interface (Figure S8A) observed for *hs* PCDH24 EC1-2 (simulations S1b-S1d, Table S10). Forced unbinding simulations were carried out at stretching speeds of 10 nm/ns (red), 1 nm/ns (blue), and 0.1 nm/ns (green, 50 ps running average in light green). (B) Snapshots of unbinding trajectory during stretching simulation at 0.1 nm/ns (S1d, Table S10). Springs indicate position and direction of applied forces. Top panel shows complex at the beginning of the simulation, other panels show snapshots at three time points indicated with black arrow heads in (A). End-to-end distance indicates the distance between the stretched atoms on the C-termini of EC2.

Figure S10. Crystal contacts in the *mm* PCDH24 EC1-3 structure. (A) Crystal contacts show a potential *cis* trimer formed by three parallel monomers. The interface area between two monomers is 1352.6 Å². Two monomers are shown as ribbons while the third one is shown in molecular surface representation. (B) Detail of trimer seen from EC1 (top) shows that the extended N-terminal strands interlace between monomers. (C) Crystal contacts show an antiparallel *trans* dimer with an interface area of 1221.2 Å². (D) Taken together, the potential *cis* and *trans* interfaces form a large complex with two *cis* trimers forming an antiparallel *trans* dimer.

Figure S11. Sequence conservation of PCDH24 EC1-3. (A) Surface representation of *mm* PCDH24 EC1-3 structure with residues colored according to sequence conservation determined using Consurf and a sequence alignment including over 94 species. Teal colors indicate residues that are least conserved while magenta indicates residues that are most conserved among species. (B) Transparent surface representation of *mm* PCDH24 EC1-3 with most conserved residues shown as an opaque magenta surface. Protein core is conserved.

Figure S12. Sequence alignment of *hs* and *mm* CDHR5 EC repeats. All 4 EC repeats for each species are aligned to each other (EC1-4).

Figure S13. Crystal contacts in the *hs* CDHR5 EC1-2 structure. (A-F) Ribbon diagram of two protomers showing various crystal contacts between monomers of *hs* CDHR5 EC1-2. Interface areas are 1433.9 Å² (A), 530.2 Å² (B), 364.7 Å² (C), 347.9 Å² (D), 132.7 Å² (E), and 114.0 Å² (F), respectively. Interfaces in (A) and (B) correspond to possible *trans* overlaps of EC1-2 and EC1-3, respectively. While *hs* CDHR5 does not mediate *trans* homophilic adhesion based on the homophilic binding assay data presented in Figure 4, these interfaces may serve as templates for the mouse CDHR5 protein. Crystal contacts in (D-F) are unlikely to be of physiological relevance due to the arrangement of the monomers.

Figure S14. Sequence conservation of CDHR5 EC1-2. (A) Surface representation of *hs* CDHR5 EC1-2 structure with residues colored according to sequence conservation determined using Consurf and a sequence alignment including over 69 species. Teal colors indicate residues that are least conserved while magenta indicates residues that are most conserved among species. (B) Transparent surface representation of *hs* CDHR5 EC1-2 with most conserved residues shown as an opaque magenta surface.

Figure S15. Homophilic binding assays of *hs* PCDH24 at various time points. (A-H) Protein G beads coated with the full-length extracellular domain of *hs* PCDH24 (A) and its C-terminal truncation versions (B-H). Images show bead aggregation observed at the start of the experiment (T0), after 30 min (T30), after 60 min (T60) followed by rocking for 1 min (R1), all in the presence of 2 mM CaCl_2 . Bar – 500 μm. (I) Protein G beads coated with the full-length extracellular domain of *hs* PCDH24 in the presence of 2 mM EDTA, shown as in (A).

Figure S16. Homophilic binding assays of *mm* CDHR5 at various time points. (A-D) Protein G beads coated with the full-length cadherin extracellular domain of *mm* CDHR5 (A) and its C-terminal truncation versions (B-D). Images show the aggregation observed at the start of experiment (T0), after 60 min (T60) followed by rocking for 1 min (R1) and 2 min (R2) in the presence of 2 mM CaCl₂. Bar – 500 μm. (E) Protein G beads coated with the full-length cadherin extracellular domain of *mm* CDHR5 in the presence of 2 mM EDTA shown as in (A).

Figure S17. Heterophilic binding assays of *hs* and *mm* PCDH24 and CDHR5. (A-B) Images from binding assays of *hs* PCDH24 (full-length extracellular domain and EC1-4Fc) mixed with the full-length *hs* CDHR5 cadherin extracellular domain (A), and *mm* PCDH24 (full-length extracellular domain and EC1-4Fc) mixed with full-length *mm* CDHR5 cadherin extracellular domain (B). Green fluorescent protein A beads are coated with PCDH24 fragments and red fluorescent protein A beads are coated with CDHR5 in all panels. The white boxes indicate the regions shown in Figure 6. Images show bead aggregation observed after 60 min followed by rocking for 2 min in the presence of 2 mM CaCl_2 . (C-F) Images from binding assays of truncations of *hs* CDHR5 mixed with *hs* PCDH24 EC1-4Fc (C), truncations of *hs* PCDH24 mixed with *hs* CDHR5 EC1-4Fc (D), truncations of *mm* CDHR5 mixed with *mm* PCDH24 EC1-4Fc (E), and truncations of *mm* PCDH24 mixed with *mm* CDHR5 EC1-4Fc (F). Images show bead aggregation observed after 60 min followed by rocking for 2 min in the presence of 2 mM CaCl_2 . (G) Images from binding assays of minimum EC repeats required for heterophilic adhesion of *hs* and *mm* PCDH24 and CDHR5. The minimum units for heterophilic adhesion for the human proteins are CDHR5 EC1Fc and PCDH24 EC1-2Fc. The minimum units for heterophilic adhesion for the mouse proteins are CDHR5 EC1-2Fc and PCDH24 EC1-2Fc. Images show bead aggregation observed after 60 minutes followed by rocking for 2 minutes in the presence of 2 mM CaCl_2 . (H) Protein A beads coated with full-length *hs* and *mm* PCDH24 and CDHR5 cadherin extracellular domains (including MAD10 for PCDH24 when indicated and without the mucin-like domain for CDHR5) in the presence of 2 mM EDTA shown as in (A-B). Bar – 500 μ m. (I) Aggregate size for minimum units of heterophilic adhesion of *hs* and *mm* PCDH24 and CDHR5 at the start of the experiment (T0), after 60 min (T60) followed by rocking for 1 min (R1) and 2 min (R2). (J) Aggregate size for full-length *hs* and *mm* PCDH24 and CDHR5 at the start of the experiment (T0), after 60 min (T60) followed by rocking for 1 min (R1) and 2 min (R2). Error bars in I and J are standard error of the mean (n indicated in Table S7).

1 Table S1. Percent Identity of EC repeats of CDH23, PCDH15, PCDH24 and CDHR5 proteins across species. The accession numbers
2 of the species used are listed in Tables S2-5.

CDH23	Repeat % identity	PCDH15	Repeat % identity	PCDH24	Repeat % identity	CDHR5	Repeat % identity
1	64.8	1	59	1	14.2	1	10.3
2	63.5	2	54.2	2	14	2	14.4
3	69.6	3	50	3	20.4	3	16.4
4	62.5	4	39.7	4	22.3	4	2.8
5	50.5	5	44.3	5	19.8	Avg	10.975
6	43.6	6	47.1	6	20.7		
7	67.3	7	37.3	7	15.8		
8	50	8	31.8	8	16.4		
9	44.8	9	39.4	9	9.8		
10	57.9	10	37.6	MAD10	10.9		
11	50.9	11	47.7	Avg	16.43		
12	55.2	MAD12	58.3				
13	43.8	Avg	45.533				
14	36.7						
15	50.5						
16	42.7						
17	51.4						
18	46.3						
19	36.4						
20	34.9						
21	42.9						
22	51.4						
23	53.3						
24	51						
25	52.7						
26	47.7						
27	53						
MAD28	69.2						
Avg	51.589						

1 Table S2. Accession numbers of PCDH24 sequences used for alignment of EC repeats across different species.

Abbreviation	Accession Number	Species	Common Name
<i>Hs</i>	NP_001165447.1	<i>Homo sapiens</i>	Human
<i>Pt</i>	XP_016809791.1	<i>Pan troglodytes</i>	Chimpanzee
<i>Rn</i>	XP_214434.4	<i>Rattus norvegicus</i>	Norwegian rat
<i>Mm</i>	NP_001028536.2	<i>Mus musculus</i>	Mouse
<i>Clf</i>	XP_005619238.1	<i>Canis lupus familiaris</i>	Dog
<i>Fc</i>	XP_019693959.2	<i>Felis catus</i>	Domestic cat
<i>Aj</i>	XP_014928203.2	<i>Acinonyx jubatus</i>	Cheetah
<i>Bt</i>	NP_001179233.2	<i>Bos taurus</i>	Cow
<i>Ss</i>	XP_013850328.2	<i>Sus scrofa</i>	Pig
<i>Oa</i>	XP_004008751.3	<i>Ovis aries</i>	Sheep
<i>Pc</i>	XP_020854166.1	<i>Phascolarctos cinereus</i>	Koala
<i>Dl</i>	XP_022420063.1	<i>Delphinapterus leucas</i>	Beluga whale
<i>Gg</i>	XP_015149366.1	<i>Gallus gallus</i>	Chicken
<i>Ccc</i>	XP_019141545.2	<i>Corvus cornix cornix</i>	Crow
<i>Af</i>	XP_019329391.1	<i>Aptenodytes forsteri</i>	Emperor penguin
<i>Ac</i>	XP_008119531.1	<i>Anolis carolinensis</i>	Green anole
<i>Dr</i>	XP_017214654.2	<i>Danio rerio</i>	Zebrafish
<i>Am</i>	XP_022532075.1	<i>Astyanax mexicanus</i>	Mexican tetra
<i>Om</i>	XP_021417551.1	<i>Oncorhynchus mykiss</i>	Rainbow trout

2

1 Table S3. Accession numbers of CDHR5 sequences used for alignment of EC repeats across different species.

Abbreviation	Accession Number	Species	Common Name
<i>Hs</i>	NP_068743.2	<i>Homo sapiens</i>	Human
<i>Pt</i>	XP_016775501.1	<i>Pan troglodytes</i>	Chimpanzee
<i>Rn</i>	XP_006230571.1	<i>Rattus norvegicus</i>	Norway rat
<i>Mm</i>	NP_001107794.1	<i>Mus musculus</i>	Mouse
<i>Fc</i>	XP_019668438.1	<i>Felis catus</i>	Domestic cat
<i>Bt</i>	NP_001096796.1	<i>Bos taurus</i>	Cow
<i>Ss</i>	XP_013845324.2	<i>Sus scrofa</i>	Pig
<i>Oa</i>	XP_027815739.1	<i>Ovis aries</i>	Sheep
<i>Pc</i>	XP_020835078.1	<i>Phascolarctos cinereus</i>	Koala
<i>Dl</i>	XP_022421100.1	<i>Delphinapterus leucas</i>	Beluga whale
<i>Af</i>	XP_009281978.1	<i>Aptenodytes forsteri</i>	Emperor penguin
<i>Ac</i>	XP_008106729.1	<i>Anolis carolinensis</i>	Green anole
<i>Dr</i>	XP_021326278.1	<i>Danio rerio</i>	Zebrafish
<i>Am</i>	XP_022525707.1	<i>Astyanax mexicanus</i>	Mexican tetra
<i>Om</i>	XP_021454527.1	<i>Oncorhynchus mykiss</i>	Rainbow trout

2

3

1 Table S4. Accession numbers of PCDH15 sequences used for EC alignment and conservation analysis.

Accession Number	Species	Common Name
NP_001136235.1	<i>Homo sapiens</i>	Human
NP_001136241.1	<i>Homo sapiens</i>	Human
XP_019792025.1	<i>Tursiops truncatus</i>	Bottlenose dolphin
XP_020929194.1	<i>Sus scrofa</i>	pig
XP_022266127.1	<i>Canis lupus familiaris</i>	Dog
NP_075604.2	<i>Mus musculus</i>	mouse
XP_024416857.1	<i>Desmodus rotundus</i>	Common vampire bat
XP_015143564.1	<i>Gallus gallus</i>	Chicken
XP_021147002.1	<i>Columbia livia</i>	Pigeon
XP_021391314.1	<i>Lonchura striata domestica</i>	Finch
XP_017589248.1	<i>Corvus brachyrhynchos</i>	American Crow
XP_020662251.1	<i>Pogona vitticeps</i>	Bearded dragon
XP_016851436.1	<i>Anolis carolinensis</i>	Green anole
XP_019394135.1	<i>Crocodylus porosus</i>	Australian crocodile
NP_001012500.1	<i>Danio rerio</i>	Zebrafish
XP_012675462.1	<i>Clupea harengus</i>	Atlantic herring
XP_022525074.1	<i>Astyanax mexicanus</i>	Mexican tetra
XP_007897895.1	<i>Callorhynchus milii</i>	Australian ghostshark

2

3

1 Table S5. Accession numbers of CDH23 sequences used for EC alignment and conservation analysis.

Accession Number	Species	Common Name
NP_071407.4	<i>Homo sapiens</i>	Human
XP_016774005.2	<i>Pan troglodytes</i>	Chimpanzee
XP_001925718.2	<i>Sus scrofa</i>	Pig
NP_001178135.2	<i>Bos taurus</i>	Cow
XP_014695776.1	<i>Equus asinus</i>	Donkey
XP_023096304.1	<i>Felis catus</i>	Cat
XP_022273252.1	<i>Canis lupus familiaris</i>	Dog
XP_007454751.1	<i>Lipotes vexillifer</i>	Yangtze river dolphin
NP_446096.1	<i>Rattus norvegicus</i>	Norwegian rat
NP_075859.2	<i>Mus musculus</i>	Mouse
XP_013028252.1	<i>Anser cygnoides domesticus</i>	Domestic goose
XP_015143732.1	<i>Gallus gallus</i>	Chicken
XP_015278377.1	<i>Gekko japonicus</i>	Japanese gekko
XP_019339897.1	<i>Alligator mississippiensis</i>	American alligator
XP_007897937.1	<i>Callorhynchus milii</i>	Australian ghostshark
NP_999974.1	<i>Danio rerio</i>	Zebrafish
XP_012672424.1	<i>Clupea harengus</i>	Atlantic herring

2

3

1 Table S6. Accession numbers of sequences of PCDH24, CDHR5, CDH23, PCDH15, CDH1, and CDH2 used for generating Figure
2 1B.

	Accession Number for EC1-3					
Species	PCDH24	CDHR5	CDH23	PCDH15	CDH1	CDH2
<i>Homo sapiens</i>	NP_001165447.1	NP_068743.3	NP_071407.4	NP_001136235.1	NP_004351.1	NP_001783.2
<i>Mus Musculus</i>	NP_001028536.2	NP_001107794.1	NP_075859.2	NP_075604.2	NP_033994.1	NP_031690.3
<i>Danio rerio</i>	XP_017214654.2	XP_021326278.1	NP_999974.1	NP_001012500.1	NP_571895.1	NP_571156.2
<i>Anolis carolensis</i>	XP_008119531.1	XP_008106723.1	XP_016847668.1	XP_016851436.1	XP_008121673.2*	XP_008106822.1
<i>Gallus gallus</i>	XP_015149366.1	NA	XP_421595.4	NP_001038119.1	NP_001034347.2	NP_001001615.1

* indicates a low-quality sequence

3

4

1 Table S7. Number of biological replicates for bead aggregation assays.

Species	Protein		Timepoints	<i>n</i>
Human	PCDH24	EC1	T = 0, 30, 60; R = 1	4
		EC1-2	T = 0, 30, 60; R = 1	4
		EC1-3	T = 0, 30, 60; R = 1	3
		EC1-4	T = 0, 30, 60; R = 1	3
		EC1-5	T = 0, 30, 60; R = 1	3
		EC1-6	T = 0, 30, 60; R = 1	3
		EC1-7	T = 0, 30, 60; R = 1	3
		EC1-MAD10	T = 0, 30, 60; R = 1	4
		EC1-MAD10*	T = 0, 60; R = 1	3
Mouse	PCDH24	EC1-MAD10	T = 0, 60; R = 1	3
Human	CDHR5	EC1-4	T = 0, 60; R = 1, 2	3
Mouse	CDHR5	EC1	T = 0, 60; R = 1, 2	3
		EC1-2	T = 0, 60; R = 1, 2	3
		EC1-3	T = 0, 60; R = 1, 2	3
		EC1-4	T = 0, 60; R = 1, 2	3
		EC1-4 E84G	T = 0, 60; R = 1, 2	3
		EC1-4 R82G	T = 0, 60; R = 1, 2	3
Human	CDHR5 EC1 + PCDH24 EC1-4	CDHR5 EC1-2 + PCDH24 EC1-4	T = 0, 60; R = 1, 2	4
		CDHR5 EC1-3 + PCDH24 EC1-4	T = 0, 60; R = 1, 2	4
		CDHR5 EC1-4 + PCDH24 EC1-4	T = 0, 60; R = 1, 2	5
		CDHR5 EC1-4 + PCDH24 EC1	T = 0, 60; R = 1, 2	3
		CDHR5 EC1-4 + PCDH24 EC1-2	T = 0, 60; R = 1, 2	3
		CDHR5 EC1-4 + PCDH24 EC1-3	T = 0, 60; R = 1, 2	3
		CDHR5 EC1-4 + PCDH24 EC1-MAD10	T = 0, 60; R = 1, 2	3
		CDHR5 EC1 + PCDH24 EC1-2	T = 0, 60; R = 1, 2	3
Mouse	CDHR5 EC1 + PCDH24 EC1-4	CDHR5 EC1-2 + PCDH24 EC1-4	T = 0, 60; R = 1, 2	3
		CDHR5 EC1-3 + PCDH24 EC1-4	T = 0, 60; R = 1, 2	3
		CDHR5 EC1-4 + PCDH24 EC1-4	T = 0, 60; R = 1, 2	3
		CDHR5 EC1-4 + PCDH24 EC1	T = 0, 60; R = 1, 2	3
		CDHR5 EC1-4 + PCDH24 EC1-2	T = 0, 60; R = 1, 2	3
		CDHR5 EC1-4 + PCDH24 EC1-3	T = 0, 60; R = 1, 2	3
		CDHR5 EC1-4 + PCDH24 EC1-MAD10	T = 0, 60; R = 1, 2	3
		CDHR5 EC1-2 + PCDH24 EC1-2	T = 0, 60; R = 1, 2	3

* Used for the *mm* PCDH24 EC1-MAD10 as a control

1 Table S8. Accession numbers of sequences of PCDH24 used in Consurf.

Accession Number	Species	Common Name
NP_001165447.1	<i>Homo sapiens</i>	Human
XP_014928203.2	<i>Acinonyx jubatus</i>	Cheetah
NP_001179233.2	<i>Bos taurus</i>	Cow
XP_005619238.1	<i>Canis lupus familiaris</i>	Dog
XP_022420063.1	<i>Delphinapterus leucas</i>	Beluga whale
XP_019693959.2	<i>Felis catus</i>	Domestic cat
NP_001028536.2	<i>Mus musculus</i>	Mouse
XP_004008751.3	<i>Ovis aries</i>	Sheep
XP_016809791.1	<i>Pan troglodytes</i>	Chimpanzee
XP_214434.4	<i>Rattus norvegicus</i>	Norwegian rat
XP_013850328.2	<i>Sus scrofa</i>	Pig
XP_014996966.2	<i>Macaca mulatta</i>	Rhesus macaque
XP_015307761.1	<i>Macaca fascicularis</i>	Crab-eating macaque
XP_012974778.1	<i>Mesocricetus auratus</i>	Golden hamster
XP_004737588.1	<i>Mustela putorius furo</i>	Domestic ferret
XP_023472938.1	<i>Equus caballus</i>	Horse
XP_029075003.1	<i>Monodon monoceros</i>	Narwhal
XP_028908976.1	<i>Ornithorhynchus anatinus</i>	Platypus
XP_008253668.2	<i>Oryctolagus cuniculus</i>	Rabbit
XP_008988594.1	<i>Callithrix jacchus</i>	White-tufted-ear marmoset
XP_024102684.1	<i>Pongo abelii</i>	Sumatran orangutan
XP_003473404.1	<i>Cavia porcellus</i>	Domestic guinea pig
XP_027263978.1	<i>Cricetulus griseus</i>	Chinese hamster
XP_023371131.1	<i>Otolemur garnettii</i>	Small-eared galago
XP_003806887.1	<i>Pan paniscus</i>	Pygmy chimpanzee
XP_003900612.1	<i>Papio anubis</i>	Olive baboon
XP_018883422.2	<i>Gorilla gorilla gorilla</i>	Western lowland gorilla
XP_004284878.1	<i>Orcinus orca</i>	Killer Whale
XP_019798890.1	<i>Tursiops truncatus</i>	Common bottlenose dolphin
XP_019329391.1	<i>Aptenodytes forsteri</i>	Emperor penguin
XP_019141545.2	<i>Corvus cornix cornix</i>	Crow
XP_015149366.1	<i>Gallus gallus</i>	Chicken
XP_027752218.1	<i>Empidonax traillii</i>	Willow Flycatcher
XP_027574630.1	<i>Pipra filicauda</i>	Wire-tailed manakin
XP_027538397.1	<i>Neopelma chrysocephalum</i>	Saffron-crested tyrant-manakin
XP_027499547.1	<i>Corapipo altera</i>	White-ruffed manakin
XP_026714386.1	<i>Athene cunicularia</i>	Burrowing owl
XP_025978389.1	<i>Dromaius novaehollandiae</i>	Emu
XP_025903300.1	<i>Nothoprocta perdicaria</i>	Chilean tinamou
XP_023791618.1	<i>Cyanistes caeruleus</i>	Blue tit
XP_021266449.1	<i>Numida meleagris</i>	Helmeted guineafowl
XP_017683677.1	<i>Lepidothrix coronata</i>	Blue-crowned manakin

XP_015731588.1	<i>Coturnix japonica</i>	Japanese quail
XP_015496956.1	<i>Parus major</i>	Great tit
XP_014803238.1	<i>Calidris pugnax</i>	Ruff
XP_013797945.1	<i>Apteryx australis mantelli</i>	North Island brown kiwi
XP_011569753.1	<i>Aquila chrysaetos canadensis</i>	American golden eagle
XP_019475977.1	<i>Meleagris gallopavo</i>	Turkey
XP_005053635.1	<i>Ficedula albicollis</i>	Collared flycatcher
XP_027634673.1	<i>Falco peregrinus</i>	Peregrine falcon
XP_027657949.1	<i>Falco cherrug</i>	Saker falcon
XP_026647227.1	<i>Zonotrichia albicollis</i>	White-throated sparrow
XP_021144536.1	<i>Columba livia</i>	Rock pigeon
XP_014108468.1	<i>Pseudopodoces humilis</i>	Tibetan ground-tit
XP_017583659.1	<i>Corvus brachyrhynchos</i>	American crow
XP_008939384.1	<i>Merops nubicus</i>	Carmine bee-eater
XP_009089922.2	<i>Serinus canaria</i>	Common canary
XP_009331800.1	<i>Pygoscelis adeliae</i>	Adelie penguin
XP_008119531.1	<i>Anolis carolinensis</i>	Green anole
XP_026576910.1	<i>Pseudonaja textilis</i>	Eastern brown snake
XP_026540941.1	<i>Notechis scutatus</i>	Mainland tiger snake
XP_020646672.1	<i>Pogona vitticeps</i>	Central bearded dragon
XP_019390868.1	<i>Crocodylus porosus</i>	Australian saltwater crocodile
XP_013925415.1	<i>Thamnophis sirtalis</i>	Common garter snake
XP_022532075.1	<i>Astyanax mexicanus</i>	Mexican tetra
XP_017214654.2	<i>Danio rerio</i>	Zebrafish
XP_021417551.1	<i>Oncorhynchus mykiss</i>	Rainbow trout
XP_014053985.1	<i>Salmo salar</i>	Atlantic salmon
XP_007904435.1	<i>Callorhynchus milii</i>	Elephant shark
XP_024139064.1	<i>Oryzias melastigma</i>	Indian Medaka
XP_022607721.1	<i>Seriola dumerili</i>	Greater amberjack
XP_015832622.1	<i>Nothobranchius furzeri</i>	Turquoise killifish
XP_004073569.1	<i>Oryzias latipes</i>	Japanese medaka
XP_029021466.1	<i>Betta splendens</i>	Siamese fighting fish
XP_028315581.1	<i>Gouania willdenowi</i>	Blunt-snouted clingfish
XP_028271476.1	<i>Parambassis ranga</i>	Indian glassy fish
XP_027865185.1	<i>Xiphophorus couchianus</i>	Monterrey platyfish
XP_026993994.1	<i>Tachysurus fulvidraco</i>	Yellow catfish
XP_026870202.1	<i>Electrophorus electricus</i>	Electric eel
XP_026802869.1	<i>Pangasianodon hypophthalmus</i>	Striped catfish
XP_026214145.1	<i>Anabas testudineus</i>	Climbing perch
XP_026186036.1	<i>Mastacembelus armatus</i>	Zig-zag eel
XP_026045273.1	<i>Astatotilapia calliptera</i>	Eastern happy
XP_019200886.1	<i>Oreochromis niloticus</i>	Nile tilapia

XP_014326181.1	<i>Xiphophorus maculatus</i>	Southern platyfish
XP_014187335.1	<i>Haplochromis burtoni</i>	Burton's mouthbrooder
XP_014342354.1	<i>Latimeria chalumnae</i>	Coelacanth
XP_007561596.1	<i>Poecilia formosa</i>	Amazon molly
XP_024918834.1	<i>Cynoglossus semilaevis</i>	Tongue sole
XP_010735796.2	<i>Larimichthys crocea</i>	Large yellow croaker
XP_010898483.1	<i>Esox lucius</i>	Northern pike
XP_012724017.1	<i>Fundulus heteroclitus</i>	Mummichog
XP_013870118.1	<i>Austrofundulus limnaeus</i>	Killifish
XP_014849149.1	<i>Poecilia mexicana</i>	Atlantic molly

1

2

1 Table S9. Accession numbers of sequences of CDHR5 used in Consurf.

Accession Number	Species	Common Name
NP_068743.2	<i>Homo sapiens</i>	Human
NP_001107794.1	<i>Mus musculus</i>	Mouse
XP_027815739.1	<i>Ovis aries</i>	Sheep
XP_022421100.1	<i>Delphinapterus leucas</i>	Beluga whale
XP_020835078.1	<i>Phascolarctos cinereus</i>	Koala
XP_016775501.1	<i>Pan troglodytes</i>	Chimpanzee
XP_006230571.1	<i>Rattus norvegicus</i>	Norway rat
XP_013845324.2	<i>Sus scrofa</i>	Pig
NP_001096796.1	<i>Bos taurus</i>	Cow
XP_019668438.1	<i>Felis catus</i>	Domestic cat
XP_023510521.1	<i>Equus caballus</i>	Horse
XP_014437989.1	<i>Tupaia chinensis</i>	Chinese tree shrew
XP_012973076.1	<i>Mesocricetus auratus</i>	Golden hamster
XP_029064308.1	<i>Monodon monoceros</i>	Narwhal
XP_028642741.1	<i>Grammomys surdaster</i>	African woodland thicket rat
XP_028370928.1	<i>Phyllostomus discolor</i>	Pale spear-nosed bat
XP_025847484.1	<i>Vulpes vulpes</i>	Red fox
XP_025717837.1	<i>Callorhinus ursinus</i>	Northern fur seal
XP_024430321.1	<i>Desmodus rotundus</i>	Common vampire bat
XP_020767741.1	<i>Odocoileus virginianus texanus</i>	White-tailed deer
XP_009281978.1	<i>Aptenodytes forsteri</i>	Emperor penguin
XP_025966497.1	<i>Dromaius novaehollandiae</i>	Emu
XP_025928961.1	<i>Apteryx rowi</i>	Okarito brown kiwi
XP_025900649.1	<i>Nothoprocta perdicaria</i>	Chilean tinamou
XP_023784276.1	<i>Cyanistes caeruleus</i>	Blue tit
XP_015485075.1	<i>Parus major</i>	Great tit
XP_015718593.1	<i>Coturnix japonica</i>	Japanese quail
XP_014796130.1	<i>Calidris pugnax</i>	Ruff
XP_014734396.1	<i>Sturnus vulgaris</i>	Common starling
XP_013800822.1	<i>Apteryx australis mantelli</i>	North Island brown kiwi
XP_013053829.1	<i>Anser cygnoides domesticus</i>	Domestic goose
XP_011595784.1	<i>Aquila chrysaetos canadensis</i>	American golden eagle
XP_010709142.1	<i>Meleagris gallopavo</i>	Turkey
XP_010562687.1	<i>Haliaeetus leucocephalus</i>	Bald eagle
XP_010297449.1	<i>Balearica regulorum gibbericeps</i>	East African grey crowned-crane
XP_010292632.1	<i>Phaethon lepturus</i>	White-tailed tropicbird
XP_010225820.1	<i>Tinamus guttatus</i>	White-throated tinamou
XP_010207507.1	<i>Colius striatus</i>	Speckled mousebird
XP_007907548.1	<i>Callorhynchus milii</i>	Elephant shark
XP_021454527.1	<i>Oncorhynchus mykiss</i>	Rainbow trout
XP_021326278.1	<i>Danio rerio</i>	Zebrafish
XP_022525707.1	<i>Astyanax mexicanus</i>	Mexican tetra

XP_024120964.1	<i>Oryzias melastigma</i>	Indian medaka
XP_023807528.1	<i>Oryzias latipes</i>	Japanese medaka
XP_029312546.1	<i>Cottoperca gobio</i>	Channel bull blenny
XP_027008843.1	<i>Tachysurus fulvidraco</i>	Yellow catfish
XP_026168289.1	<i>Mastacembelus armatus</i>	Zig-zag eel
XP_019112627.2	<i>Larimichthys crocea</i>	Large yello croaker
XP_017341402.1	<i>Ictalurus punctatus</i>	Channel catfish
XP_016392038.1	<i>Sinocyclocheilus rhinoceros</i>	Rayfinned fish
XP_016324417.1	<i>Sinocyclocheilus anshuiensis</i>	Rayfinned fish
XP_015227070.1	<i>Cyprinodon variegatus</i>	Sheepshead minnow
XP_021163353.1	<i>Fundulus heteroclitus</i>	Mummichog
XP_012674652.1	<i>Clupea harengus</i>	Atlantic herring
XP_005938389.1	<i>Haplochromis burtoni</i>	Burton's mouthbrooder
XP_004556350.2	<i>Maylandia zebra</i>	Zebra mbuna
XP_008106729.1	<i>Anolis carolinensis</i>	Green anole
XP_019340716.1	<i>Alligator mississippiensis</i>	American alligator
XP_028591121.1	<i>Podarcis muralis</i>	Common wall lizard
XP_026562223.1	<i>Pseudonaja textilis</i>	Eastern brown snake
XP_026523924.1	<i>Notechis scutatus</i>	Mainland tiger snake
XP_026513403.1	<i>Terrapene carolina triunguis</i>	Three-toed box turtle
XP_019402850.1	<i>Crocodylus porosus</i>	Australian saltwater crocodile
XP_019375347.1	<i>Gavialis gangeticus</i>	Gharial
XP_015745583.1	<i>Python bivittatus</i>	Burmese python
XP_013925033.1	<i>Thamnophis sirtalis</i>	Common garter snake
XP_014429985.1	<i>Pelodiscus sinensis</i>	Chinese softshell turtle
XP_025054327.1	<i>Alligator sinensis</i>	Chinese alligator
XP_023967188.1	<i>Chrysemys picta bellii</i>	Western painted turtle

1

2

1 Table S10. Summary of simulations

Label	System	Type	Start	Time (ns)	Pulling speed (nm/ns)	Average force peak (pN) ^d	Size of system (# atoms)	Box dimensions (Å ³)
S1a	<i>Hs</i> PCDH24 EC1-2	Equil ^a	-	10.8	-	-	125,348	227 × 88 × 67
S1b		SMD ^b	S1a	1	10	675.2		
S1c		SMD ^b	S1a	21	1	383.9		
S1d		SMD ^b	S1a	58.6	0.1	167.1		
S2	<i>Mm</i> PCDH24 EC1-3	Equil ^a	-	99.1	-	-	227,299	171 × 112 × 111
S3a	<i>Mm</i> PCDH24 EC1-3	Equil ^a	-	11.1	-	-	195,074	337 × 71 × 80
S3b		SMD ^c	S3a	1	10	1905.5		
S3c		SMD ^c	S3a	10	1	1385.4		
S3d		SMD ^c	S3a	70	0.1	1216.0		

2 ^a Indicates simulations that consisted of 1,000 steps of minimization, 100 ps of dynamics with the protein backbone constrained, 1 ns
3 of free dynamics in the *NpT* ensemble ($\gamma = 1 \text{ ps}^{-1}$), and then of free dynamics in the *NpT* ensemble ($\gamma = 0.1 \text{ ps}^{-1}$) for the rest of the time.

4 ^b Indicates SMD simulations in which the force was applied to the C α atoms of residues p.P215 of chains A and D.

5 ^c Indicates SMD simulations in which the force was applied to the centers of mass (COM) of EC1 (p.N1 to p.E95) and EC3 (p.P215
6 and p.T324).

7 ^d Average peak forces were calculated from peak forces applied to stretched C α atoms or COM at each end and using 50 ps running
8 averages.

9

10

11

12