PNAS
www.pnas.org

Supplementary Information for

Interaction specificity of clustered protocadherins inferred from sequence covariation and structural analysis

John M. Nicoludis[a,b,1], Anna G. Green[c,1], Sanket Walujkar[d], Elizabeth J. May[a], Marcos Sotomayor[d], Debora S. Marks[c,2], Rachelle Gaudet[a,2]

[a]Department of Molecular and Cellular Biology, Harvard University, Cambridge MA, USA
[b]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA
[c]Department of Systems Biology, Harvard Medical School, Boston, MA, USA
[d]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA

[1]These authors contributed equally to this work
[2]Corresponding authors: debbie@hms.harvard.edu, gaudet@mcb.harvard.edu

This PDF file includes:

Supplementary text
Figs. S1 to S9
Tables S1 to S4
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S3

www.pnas.org/cgi/doi/10.1073/pnas.1821063116

Supplementary Information Text

Statistical model explains changes in Pcdh specificity in chimera mutants. The statistical energy of interaction model allows us to predict how particular mutations may alter interaction specificity by recalculating SEI using coupling terms from the mutant sequence. Previous work has tested chimeric constructs in cell aggregation assays to understand how specificity is encoded in the clustered Pcdh family (1). In these experiments, chimeras are constructed by combining domains from a parent isoform and a target isoform. The original chimeras did not interact with the parent or target. Residues in the parent domains were mutated to residues found in the target domain, and the resulting mutant chimeras interact with the target but not the parent.

To validate our model on these experimental data, we calculated the SEI between the chimera and target isoforms, and the mutant chimera and target isoforms (**Fig. S13A**). If the change in SEI from the chimera to the mutant chimera (ΔSEI) is positive, it means the mutant chimera is more likely to interact with the target isoform than the original chimera. If ΔSEI is negative, the mutant chimera would be less likely to interact with the target isoform than the original chimera. In Rubinstein et al. (2015), three pairs of closely related isoforms (>85% identity) were chosen: Pcdhα7/Pcdhα8, PcdhγA8/PcdhγA9 and Pcdhβ6/ Pcdhβ8 (1). Out of the seven mutant chimeras that were able to interact with the target isoform, we correctly predicted four, observing that the introduced mutations result in an increase in SEI (**Fig. S13B**). The three predicted incorrectly lead to only a very small change in SEI. Overall our statistical energy model of interaction specificity, when enough high-quality sequence information is available, is very consistent with available experimental data.

The experiments by Goodman and colleagues were performed on closely related isoforms, and therefore only a small number of possible mutations could be tested for their ability to reprogram specificity. If future experiments seek to reprogram more distantly related isoforms, models such as ours will be needed to inform the choice of residues to mutate.

Experimental Methods

Further explanation of statistical interaction energy model. To determine the single mutations most likely to reprogram clustered Pcdh A to interact with clustered Pcdh B, we first calculated the SEI of A with A and of A with B. We then computationally swap in each residue, one at a time, from B into A, and assess SEI of this new A* chimera with sequence A and with sequence B. We compute a change in energy (ΔSEI) between this mutant sequence and the wild type SEI values. We excluded the C-type isoforms from this analysis because they are evolutionarily distinct and have unique biological functions (2–4). Model parameters are available on request.

Crystallography of PcdhγB3 EC1-4 without HEPES. PcdhγB3 EC1-4 was purified as previously reported (5) and crystallized in 10% PEG 5000 monomethylester and 4% ethylene glycol with either 50 mM HEPES pH 7 or 100 mM Tris pH 7. Crystals were cryo-protected in reservoir plus 20% glycerol and cryocooled in liquid N_2. X-ray data were collected according to **Table S3** and processed in HKL2000 (6). Structures were determined using molecular replacement and the published PcdhγB3 EC1-4 structure (5). Model building was done in COOT (7). Refinement and generation of composite omit maps was done in PHENIX (8). Software is maintained through SBGrid (9). Refinement and model statistics are listed in **Table S3.**

MD simulations. Four different crystal structures of clustered Pcdhs – mouse Pcdhα7 (5dzv), β6 (5dzx), γB7 (5szp) and human PcdhγB3 (5k8r) were each solvated using TIP3P water (**Table S1**). All the systems were neutralized and ionized with 150 mM NaCl. Resulting systems were minimized for 5,000 steps and equilibrated for 200 ps with backbone constraints ($k = 1$ kcal/mol/Å^2)

and for an additional 1 ns without these constraints. During these initial 1.2 ns of simulation a Langevin damping coefficient of $\gamma = 1.0$ ps^{-1} was used. Subsequent dynamics used $\gamma = 0.1$ ps^{-1}. All these simulations were performed using NAMD 2.12, the CHARMM36/CMAP force field (10, 11), and periodic boundary conditions in the *NpT* ensemble with $T = 300$ K and $p = 1$ atm. Each system was simulated for a total of 120 ns. We used a uniform integration time step of 2 fs with SHAKE, the particle mesh Ewald method for computation of long-range electrostatic interactions (grid point density > 1 Å^3), and a cutoff radius for van der Waals interactions of 12 Å. For PcdhγB3, periodic images of the protein briefly came within less than ~12 Å of each other and quickly moved away. We do not believe this affected the dynamics of the protein adversely.

Buried surface area (BSA) was calculated every 10 ps for each trajectory using the "measure sasa" tool in VMD (12) with 1.4 Å sampling radius. We first computed the solvent accessible surface area (SASA) of individual protomers, and then the SASA of the complex formed by that pair. The final BSA reported was calculated by subtracting the SASA of the complex from the addition of the SASAs of each individual subunit.

Defining interface residues using structural analysis and MD simulations. To build a model for Pcdh interaction specificity, we needed to define a set of residues that constitute the interface for clustered Pcdhs. First, we determined the interface residues (defined as closest heavy atoms are within 8 Å) from all available crystal structures of validated antiparallel clustered Pcdh interfaces: mouse PcdhγA1 EC1-3 (4zi9), mouse Pcdhα7 EC1-5 (5dzv), mouse Pcdhα4 EC1-4 (5dzw), mouse Pcdhβ6 EC1-4 (5dzx), mouse Pcdhβ8 EC1-4 (5dzy), human PcdhγB3 EC1-4 (5k8r), mouse PcdhγA8 EC1-4 (5szm), mouse PcdhγB7 EC1-4 (5szp), and mouse PcdhγB2 EC1-5 (5t9t) (5, 13–15). This collection of structures yielded a structure-based set of 205 total residues with 1634 residue pairs using the 8 Å cutoff (**Table S4).**

Second, we analyzed our MD trajectories to define a simulations-based set of interacting residues, defining an interacting residue pair as a pair of amino acids with non-hydrogen atoms that come within 5 Å of each other in at 10% of the simulation frames sampled at 100 ps. From this definition, the interface is defined by a total of 1880 residue pairs originating from a total of 236 interface residues using the 10% cutoff **(Table S4, Fig. S14)**. Of the 205 interface residues determined from crystal structures alone, 204 are also in the 236 interface residues from the MD simulations, reflecting agreement between these two definitions of clustered Pcdh interface residues. The somewhat expanded simulations-based set likely includes residues that transiently participate in the interface and/or participate in the interface in other isoforms. We thus used this set of 236 interface residues for our sequence-based analyses (**SI Dataset 1**).

Construction of sequence alignment. Sequences were found by aligning the PcdhγB3 isoform (Uniprot identifier: PCGDF_HUMAN) against the Uniref database (download date: April 2016) using Jackhmmer with 5 iterations (16). The alignment was filtered to contain only clustered Pcdh sequences, as described in our earlier work (5, 13). The alignment was renumbered according to the mouse Pcdα7 isoform (Uniprot identifier: PCDA7_MOUSE), and gaps were defined relative to this sequence. The alignment was filtered to remove sequences that contain more than 50% gaps and to remove columns that contain more than 50% gaps. The alignment was truncated to contain only domains EC1 through EC4. This alignment has 8560 sequences, with effective number of sequences (M_{eff}) of 3300 after down-weighting sequences that are more than 90% identical.

Iterative pairing algorithm. Previous work has shown that interacting paralogs from bacteria can be correctly matched by iteratively building a sequence alignment with pairs that have the best SEI (17, 18). We use a similar approach for the Pcdhs, where we seed an alignment with 1000 randomly paired sequences of EC1-EC2 and EC3-EC4 domains. Each EC1-EC2 is paired with randomly EC3-EC4 from the same species. For speed, we infer couplings using the mean field approximation (19, 20) as implemented in (21). We then assess the SEI of all possible pairs of sequences within the same species, and keep the top ones for the next iteration. The algorithm was run for 300 iterations, increasing the alignment size by 50 sequences per iteration. Each experiment was repeated for five replicates.

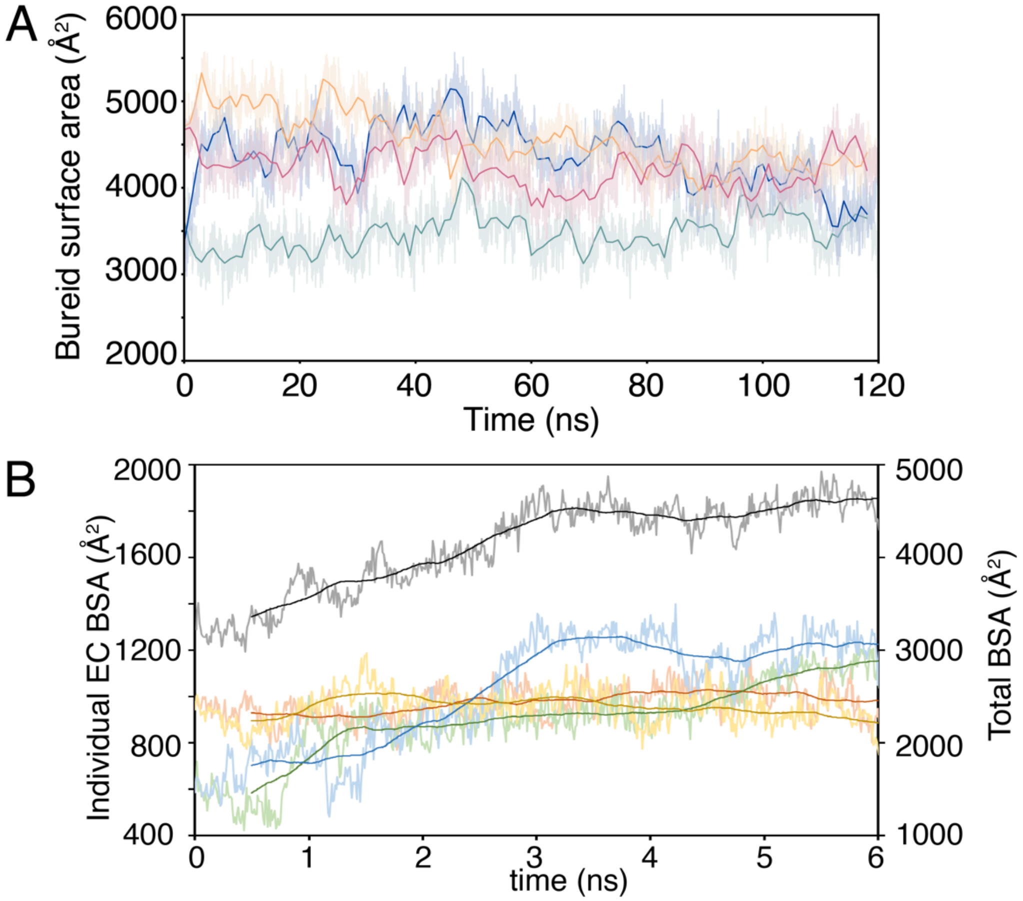
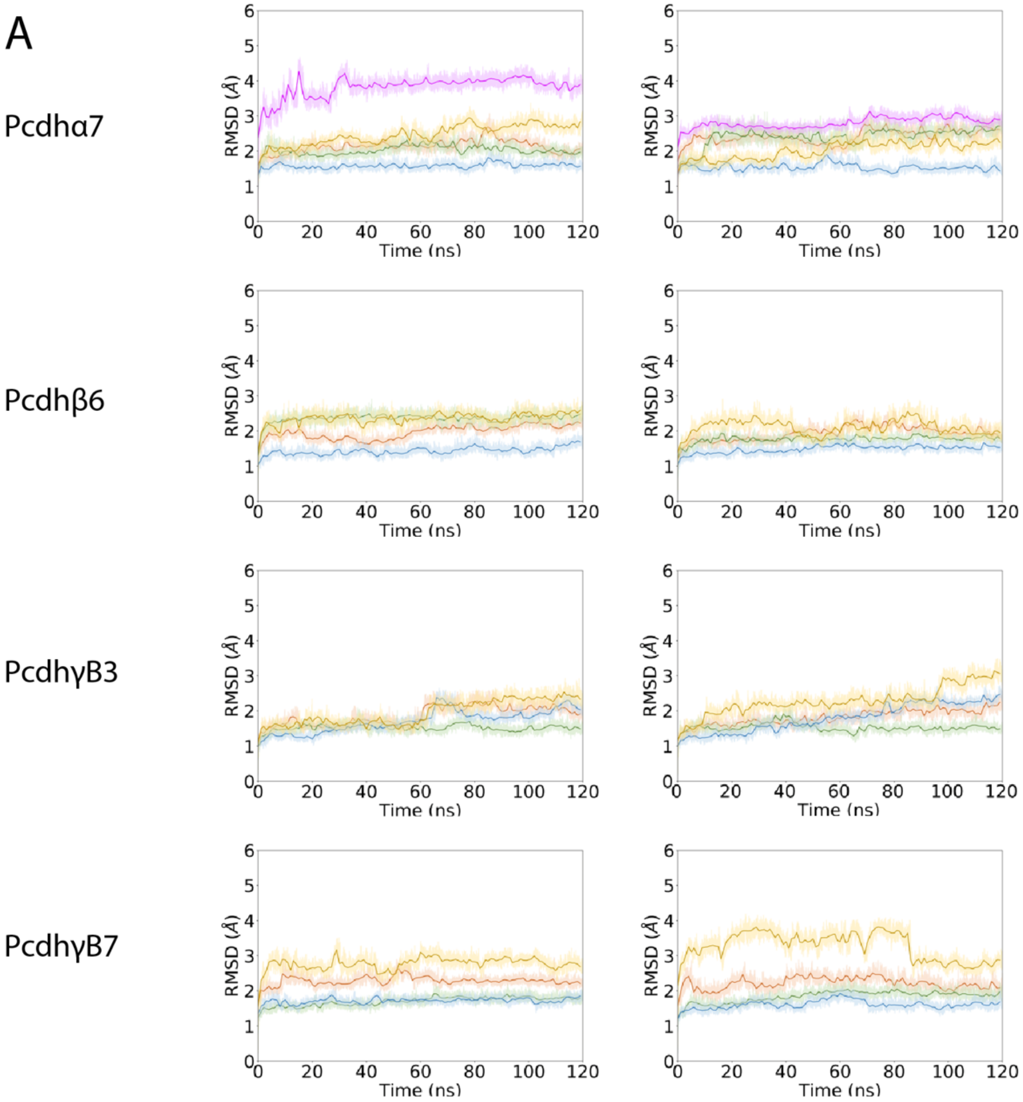


Fig. S1. Buried surface area of Pcdh dimer simulations

(A) The overall buried surface areas (BSAs) of the PcdhγB3 (dark blue), PcdhγB7 (orange), Pcdhβ6 (magenta), and Pcdhα7 (teal) complexes are large (3400-5000 Å^2) and consistent throughout the simulations. (B) BSA values for the PcdhγB3 plotted for the first 6 ns show that the overall BSA of the dimer increases from ~3000 Å^2 to 4500 Å^2 in the first 3 ns of the simulation (black). This increase in BSA is due to independent increases at the EC2/EC3 interfaces (blue and green), while the EC1/EC4 interfaces maintain the same BSA (yellow and red).


A
Pcdhα7
Pcdhβ6
PcdhγB3
PcdhγB7
RMSD (Å)
0
1
2
3
4
5
6
Time (ns)
0
20
40
60
80
100
120

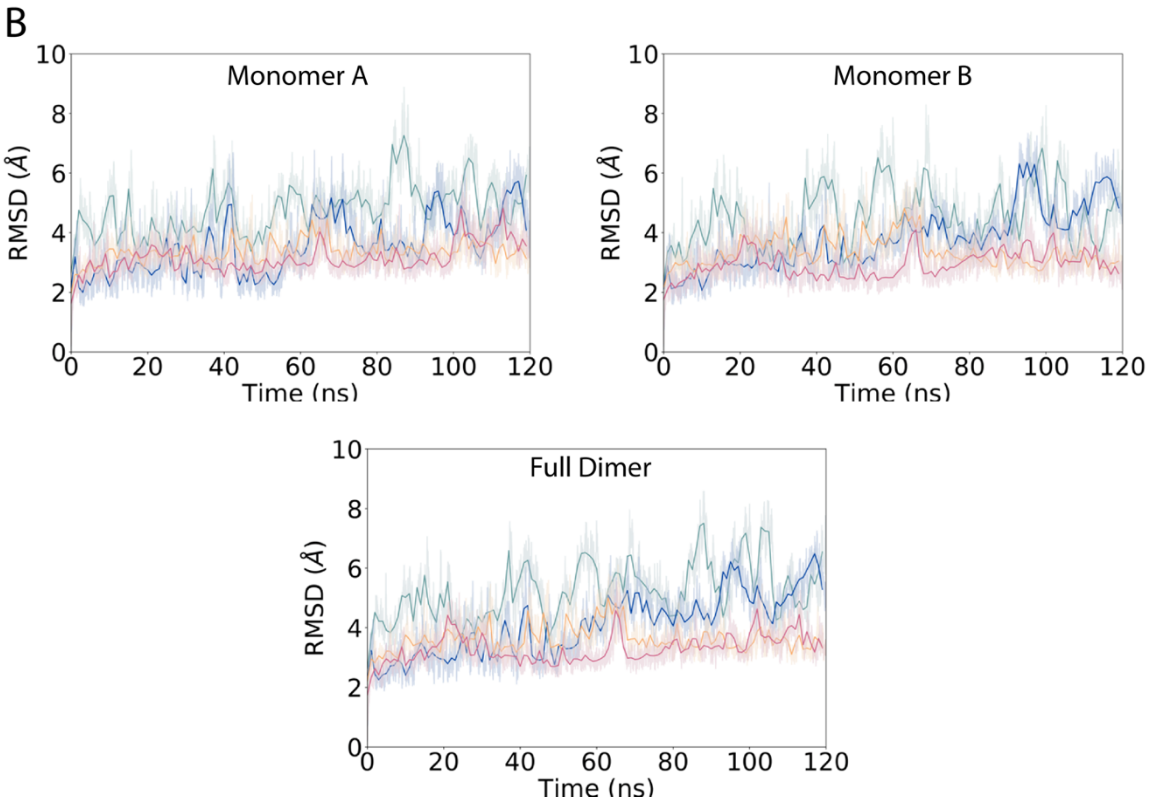


Fig. S2. Clustered protocadherin RMSD
(A) The RMSD values of individual ECs for the Pcdh dimers range from 2-3 Å. EC1 is in orange, EC2 in green, EC3 in blue, EC4 in yellow, and EC5 in purple (for Pcdhα7). The two plots represent the two protomers of the Pcdh complex. (B) RMSD of the two monomers in the dimer (PcdhγB3 (dark blue), PcdhγB7 (orange), Pcdhβ6 (magenta), and Pcdhα7 (teal)) and overall RMSD of the full dimer.

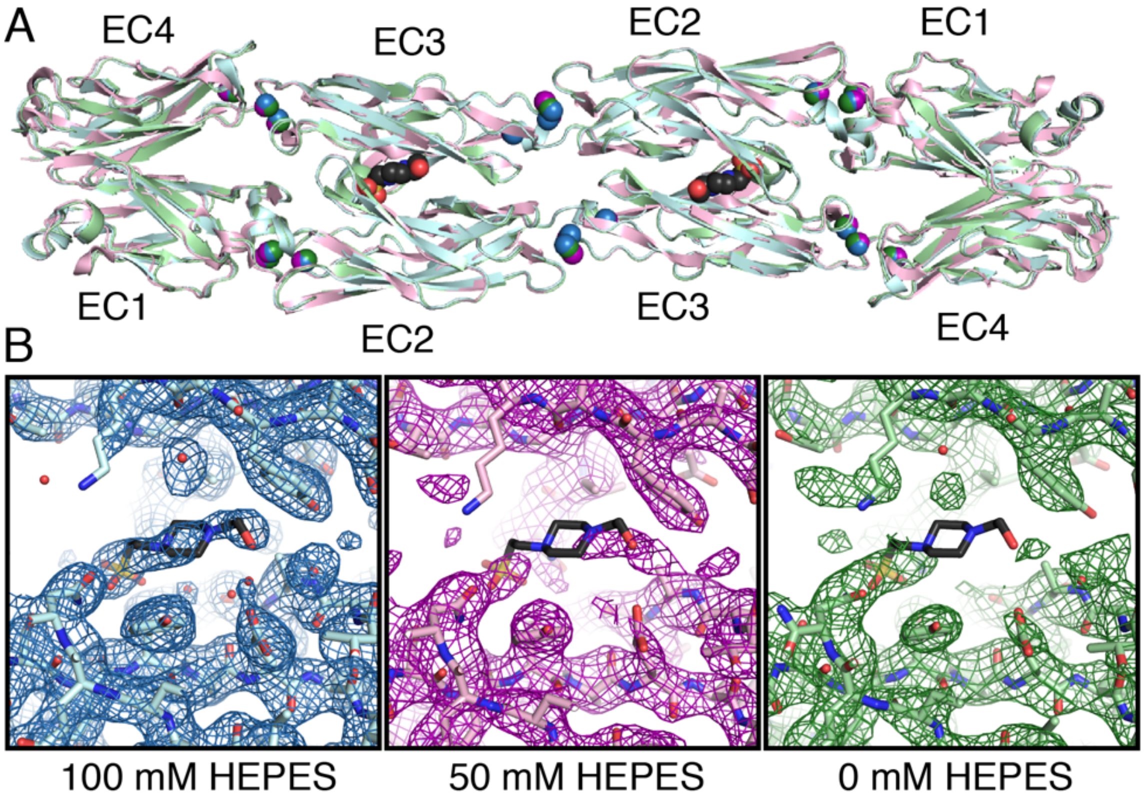


Fig. S3. HEPES does not influence the structure of PcdhγB3 EC1-4
(A) Crystal structures of PcdhγB3 EC1-4 in the presence of 100 mM HEPES (blue; PDB ID: 5k8r), 50 mM HEPES (pink; PDB ID: 6mer), and no HEPES (green; PDB ID: 6meq) adopt the same conformation. The HEPES molecule in the original structure (PDB ID: 5k8r) is shown in black for carbon, red for oxygen, blue for nitrogen, and yellow for sulfur. (B) Composite omit maps contoured at 1 σ (which reduce model bias in electron density) of these three structures show density for HEPES in the original structure, but less or no density in the lower and no HEPES conditions.

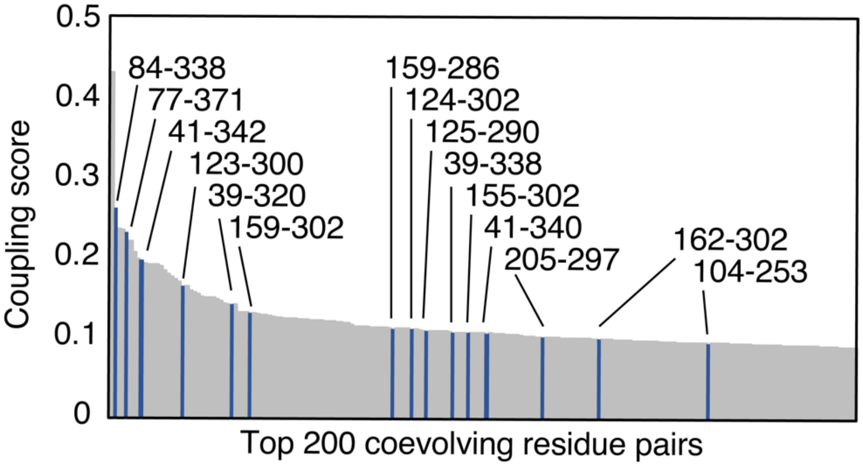


Fig. S4. Rank order of intermolecular coevolving residue pairs
Evolutionary coupling scores of the top 200 coevolving residues pairs includes the top 15 intermolecular pairs, highlighted in blue and labeled with corresponding residue numbers in PcdhγB3. All other pairs are intramolecular.

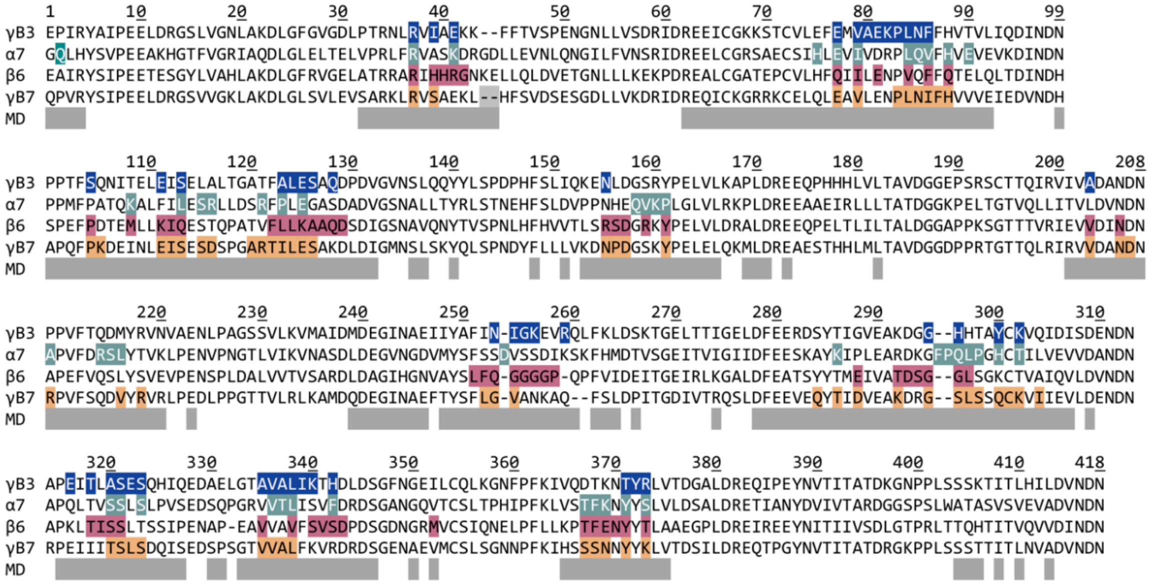

1 10 20 30 40 50 60 70 80 90 99
γB3 EPIRYAIPEELDRGSLVGNLAKDLGFGVGDLPTRNLRVIAEKK--FFTVSPENGNLLVSDRIDREEICGKKSTCVLEFEMVAEKPLNFFHVTVLIQDINDN
α7 GQLHYSVPEEAKHGTFVGRIAQDLGLELTELVPRLFRVASKDRGDLLEVNLQNGILFVNSRIDREELCGRSAECSIHLEVIVDRPLQVFHVEVEVKDINDN
β6 EAIRYSIPEETESGYLVAHLAKDLGFRVGELATRRARIHHRGNKELLQLDVETGNLLLKEKPDREALCGATEPCVLHFQIILENPVQFFQTELQLTDINDH
γB7 QPVRYSIPEELDRGSVVGKLAKDLGLSVLEVSARKLRVSAEKL--HFSVDSESGDLLVKDRIDREQICKGRRKCELQLEAVLENPLNIFHVVVEIEDVNDH
MD
110 120 130 140 150 160 170 180 190 200 208
γB3 PPTFSQNITELEISELALTGATFALESAQDPDVGVNSLQQYYLSPDPHFSLIQKENLDGSRYPELVLKAPLDREEQPHHHLVLTAVDGGEPSRSCTTQIRVIVADANDN
α7 PPMFPATQKALFILESRLLDSRFPLEGASDADVGSNALLTYRLSTNEHFSLDVPPNHEQVKPLGLVLRKPLDREEAAEIRLLLTATDGGKPELTGTVQLLITVLDVNDN
β6 SPEFPDTEMLLKIQESTQPATVFLLKAAQDSDIGSNAVQNYTVSPNLHFHVVTLSRSDGRKYPELVLDRALDREEQPELTLILTALDGGAPPKSGTTTVRIEVVDINDN
γB7 APQFPKDEINLEISESDSPGARTILESAKDLDIGMNSLSKYQLSPNDYFLLLVKDNPDGSKYPELELQKMLDREAESTHHLMLTAVDGGDPPRTGTTQLRIRVVDANDN
MD
220 230 240 250 260 270 280 290 300 310
γB3 PPVFTQDMYRVNVAENLPAGSSVLKVMAIDMDEGINAEIIYAFIN-IGKEVRQLFKLDSKTGELTTIGELDFEERDSYTIGVEAKDGG--HHTAYCKVQIDISDENDN
α7 APVFDRSLYTVKLPENVPNGTLVIKVNASDLDEGVNGDVMYSFSSDVSSDIKSKFHMDTVSGEITVIGIIDFEESKAYKIPLEARDKGFPQLPGHCTILVEVVDANDN
β6 APEFVQSLYSVEVPENSPLDALVVTVSARDLDAGIHGNVAYSLFQ-GGGGP-QPFVIDEITGEIRLKGALDFEATSYYTMEIVATDSG--GLSGKCTVAIQVLDVNDN
γB7 RPVFSQDVYRVRLPEDLPPGTTVLRLKAMDQDEGINAEFTYSFLG-VANKAQ--FSLDPITGDIVTRQSLDFEEVEQYTIDVEAKDRG--SLSSQCKVIIEVLDENDN
MD
320 330 340 350 360 370 380 390 400 410 418
γB3 APEITLASESQHIQEDAELGTAVALIKTHDLDSGFNGEILCQLKGNFPFKIVQDTKNTYRLVTDGALDREQIPEYNVTITATDKGNPPLSSSKTITLHILDVNDN
α7 APQLTVSSLSLPVSEDSQPGRVVTLISVFDRDSGANGQVTCSLTPHIPFKLVSTFKNYYSLVLDSALDRETIANYDVIVTARDGGSPSLWATASVSVEVADVNDN
β6 APKLTISSLTSSIPENAP-EAVVAVFSVSDPDSGDNGRMVCSIQNELPFLLKPTFENYYTLAAEGPLDREIREEYNITIIVSDLGTPRLTTQHTITVQVVDINDN
γB7 RPEIIITSLSDQISEDSPSGTVVALFKVRDRDSGENAEVMCSLSGNNPFKIHSSSNNYYKLVTDSILDREQTPGYNVTITATDRGKPPLSSSTTITLNVADVNDN
MD

Fig. S5. Sequence alignment of clustered protocadherin isoforms

EC1-4 amino acid sequence and residue numbering of clustered protocadherin isoforms (PcdhγB3, Pcdhα7, Pcdhβ6, and PcdhγB7) on which we performed MD simulations. Interface residues from crystal structures are shown in the respective colors and contacting residues from simulations are highlighted in gray below.

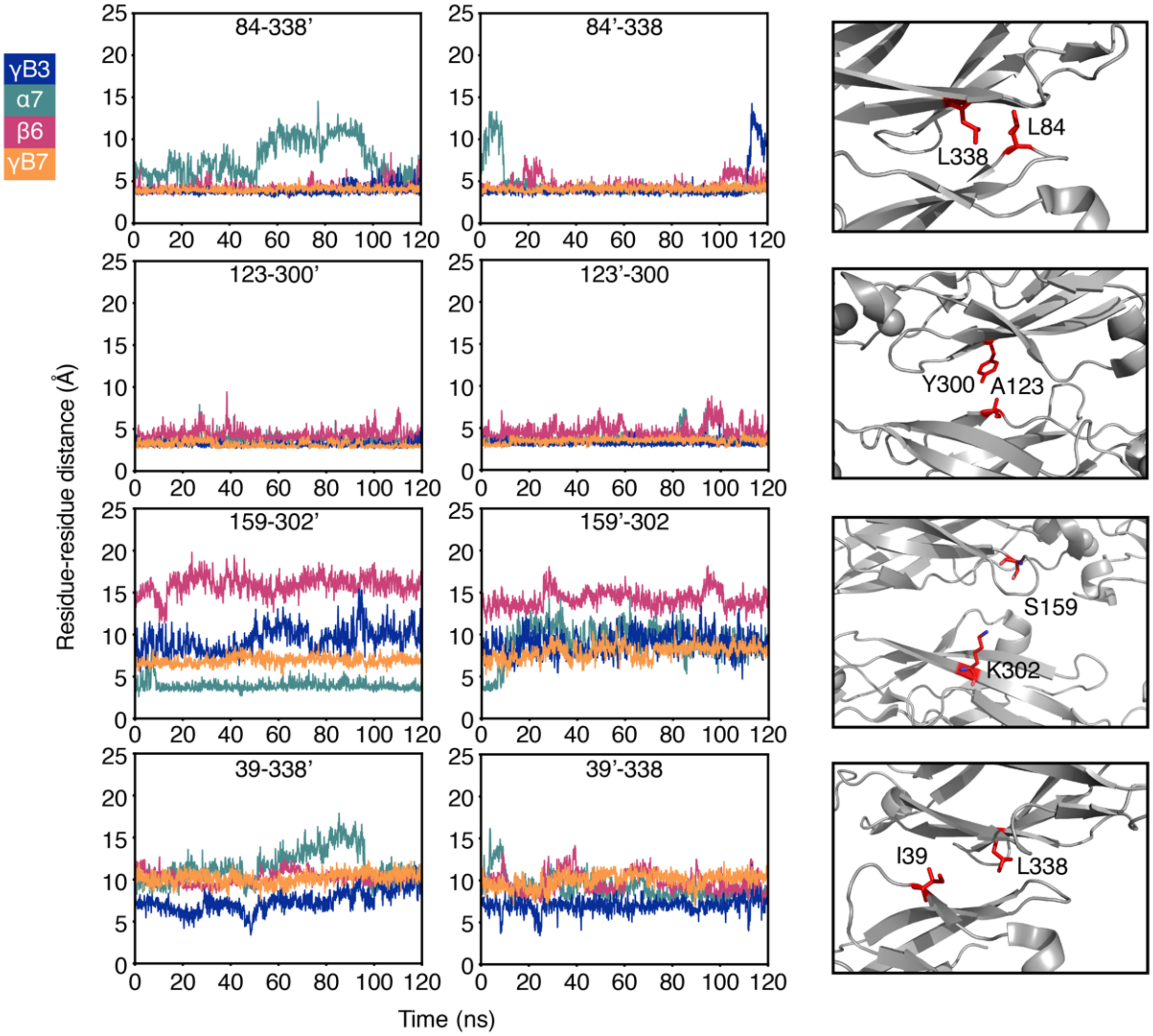


Fig. S6. Diverse residue-residue distance trajectories for coevolving residue pairs
Examples of residue-residue distances plotted over simulation time for some of the top 15 intermolecular coevolving residue pairs. Plots on the left show residue-residue distances for the four simulations (PcdhγB3 (dark blue), PcdhγB7 (orange), Pcdhβ6 (magenta), and Pcdhα7 (teal)) with the two semi-independent pairs from each dimer plotted on the left and right graph, respectively. On the right each residue pair in the Pcdh γB3 crystal structure (PDB ID: 5k8r) is illustrated in red stick representation.

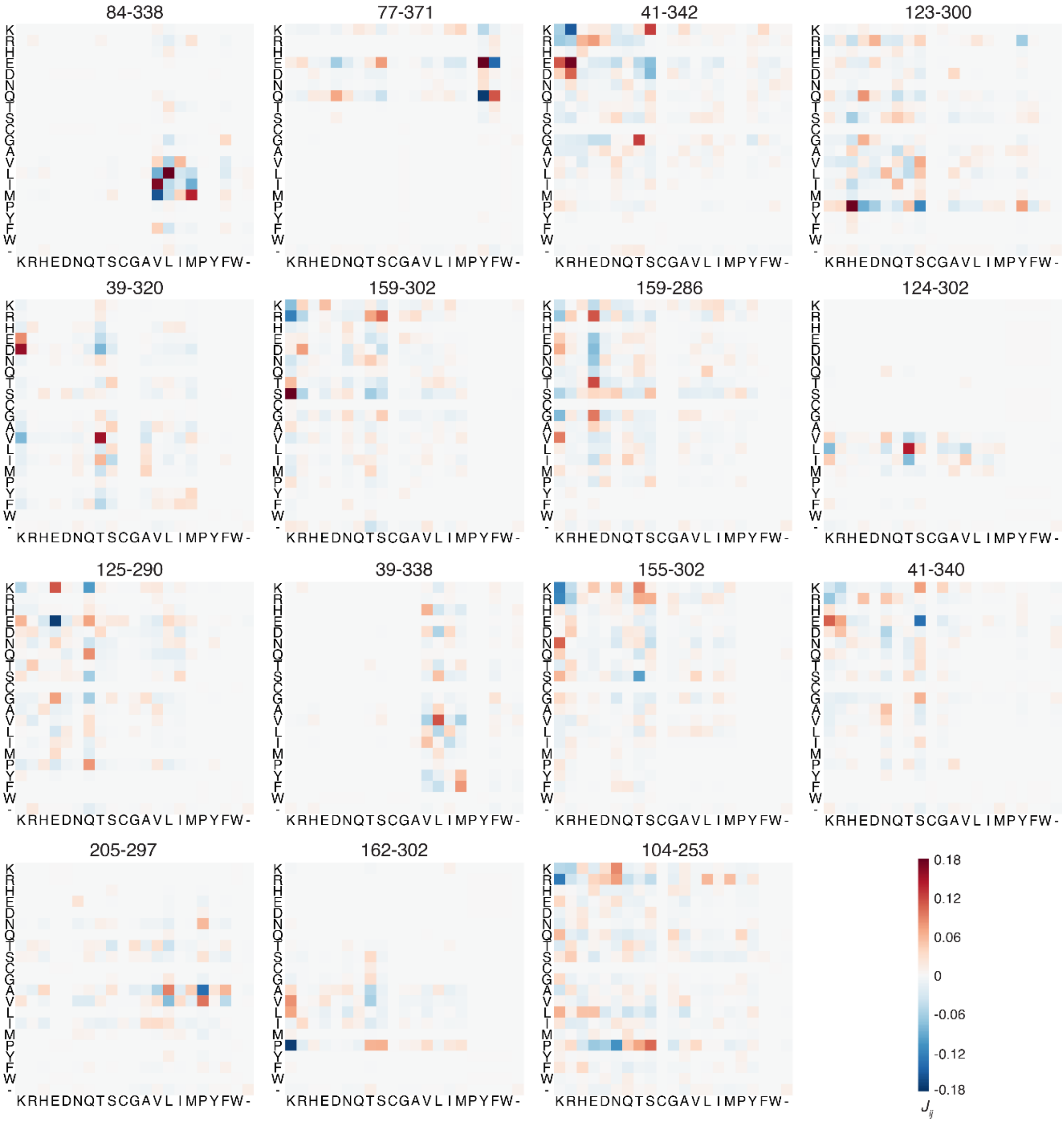


Fig. S7. Distribution of J_{ij} values for top 15 intermolecular coupled pairs.
Distributions of J_{ij} values for the top 15 intermolecular coevolving pairs are distinct for the strongly coupled pairs on EC1/EC4 (e.g. 84-338 and 77-371) versus the more weakly coupled pairs on EC2/EC3 (e.g. 123-300, 159-286). The difference in distribution reflects the biochemical similarity of strongly coupled pairs on EC1/EC4 and the diversifying selection present on EC2/EC3.

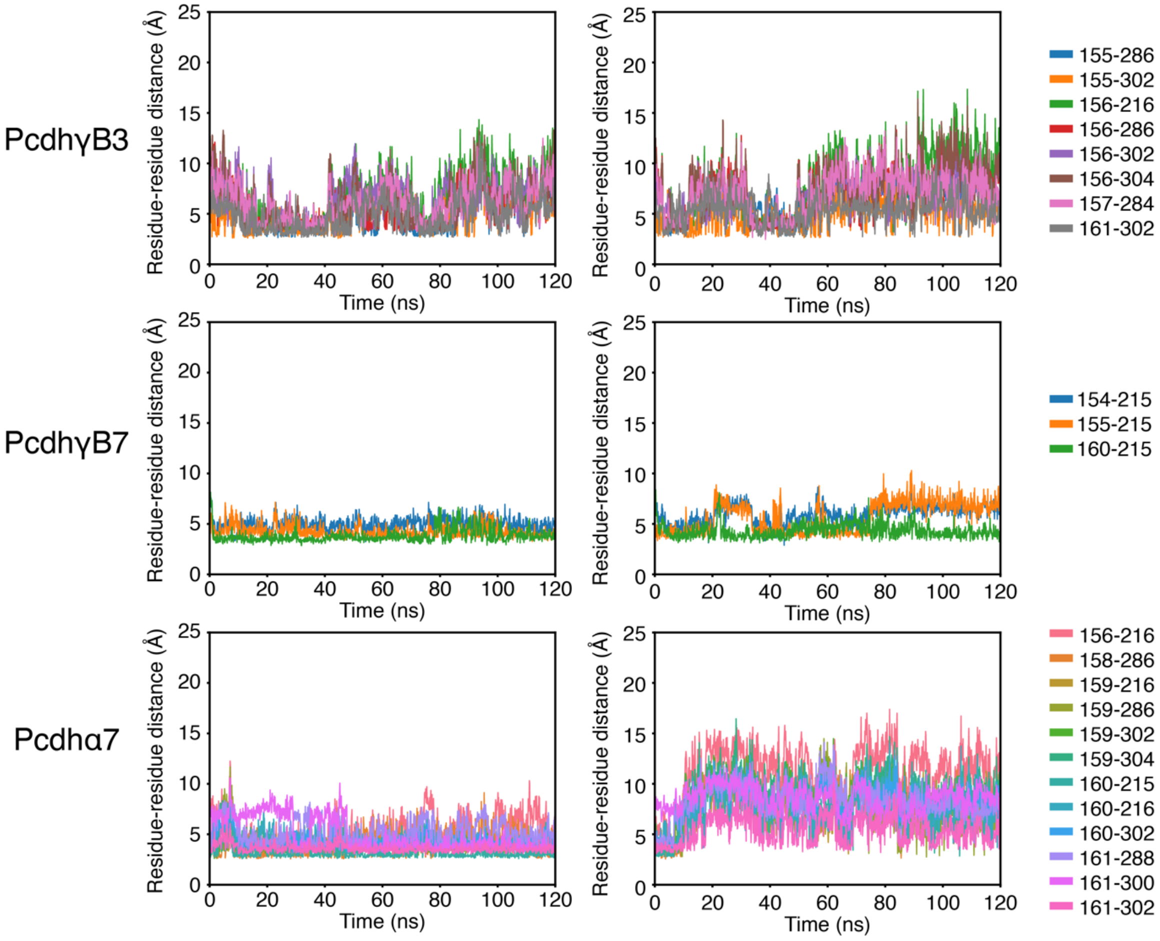


Fig. S8. Interactions between the EC2 β4-β5 loop and EC3 are dynamic in nature
Residues in the EC2 β4-β5 loop (positions 154-161) interact with residues on EC3 (positions 215, 216, 286, 288, 300, 302, and 304) dynamically as seen in the residue-residue distances of these positions throughout MD simulations. Positions shown are those that have significant contact within that simulation (see Fig. 4B). This loop in Pcdhβ6 does not make significant contact with EC3 and thus is not shown.

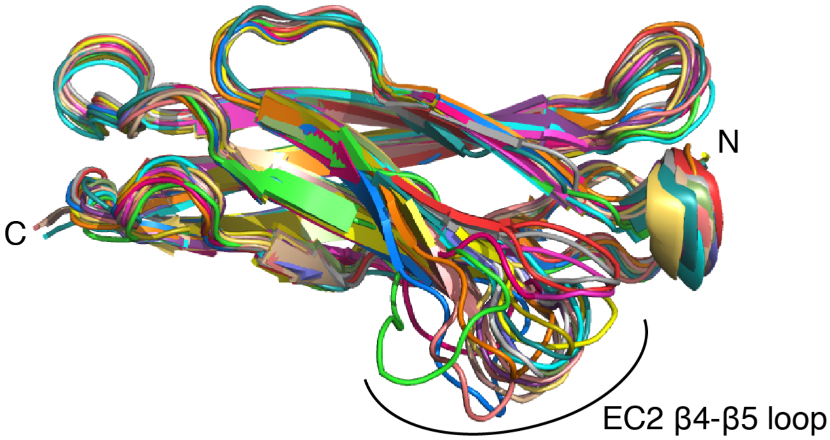


Fig. S9. Conformational diversity of EC2 β4-β5 loop in protocadherin structures
Superposition of EC2 of the following protocadherin structures shows conformational diversity of the β4-β5 loop:

lime green	Pcdhα4	5dzw
cyan	PcdhγB3	5k8r
magenta	Pcdhβ1	4zpl
yellow	Pcdhαc2	4zpm
light pink	PcdhγC5	4zpo
white	PcdhγC5	4zpp
slate blue	PcdhγC5	4zpq
orange	PcdhγA8	4zps
teal	Pcdh19	5iu9
bright pink	Pcdhα7	5dzv
light yellow	Pcdhβ6	5dzx
purple	Pcdhβ8	5dzy
gray	PcdhγA1	5szl
sky blue	PcdhγA8	5szm
gold	PcdhγA9	5szn
light green	PcdhγC3	4zi8
red	PcdhγA1	4zi9
mauve	PcdhγB7	5szp
light orange	PcdhγB7	5szo
light red	PcdhγB2	5t9t

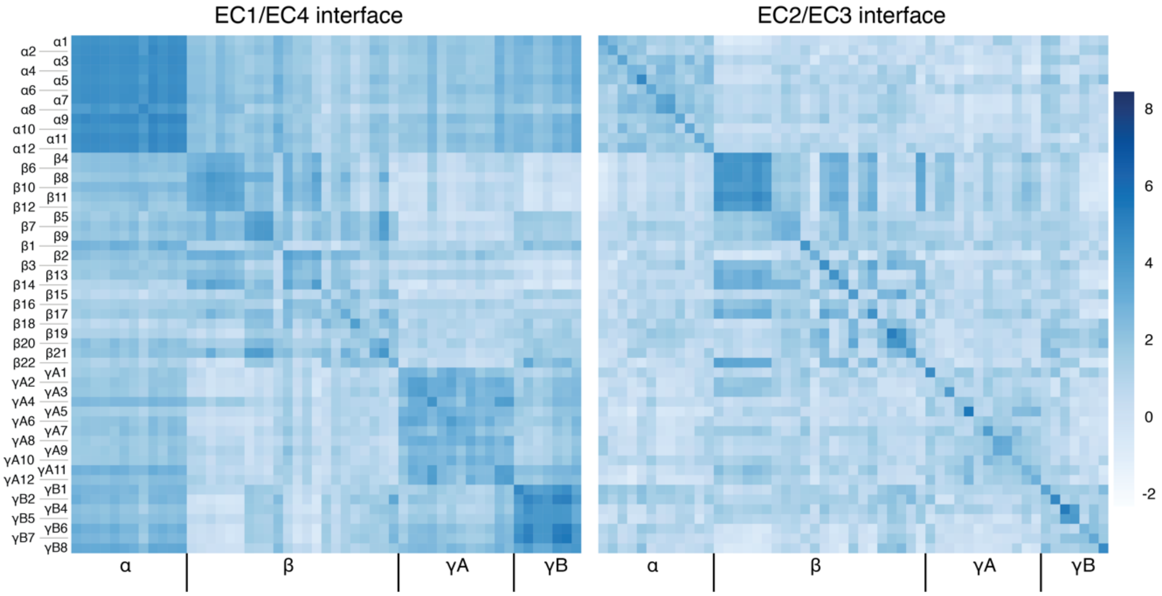


Fig. S10. EC2/EC3 interface SEI is more distinct between isoforms than EC1/EC4 interface SEI

SEI for all possible isoform pairings computed for the EC1/EC4 and EC2/EC3 interfaces. For the α, γA, and γB subfamilies in particular, the EC2/EC3 interface shows higher preference for self versus nonself interactions compared to the EC1/EC4 interface.

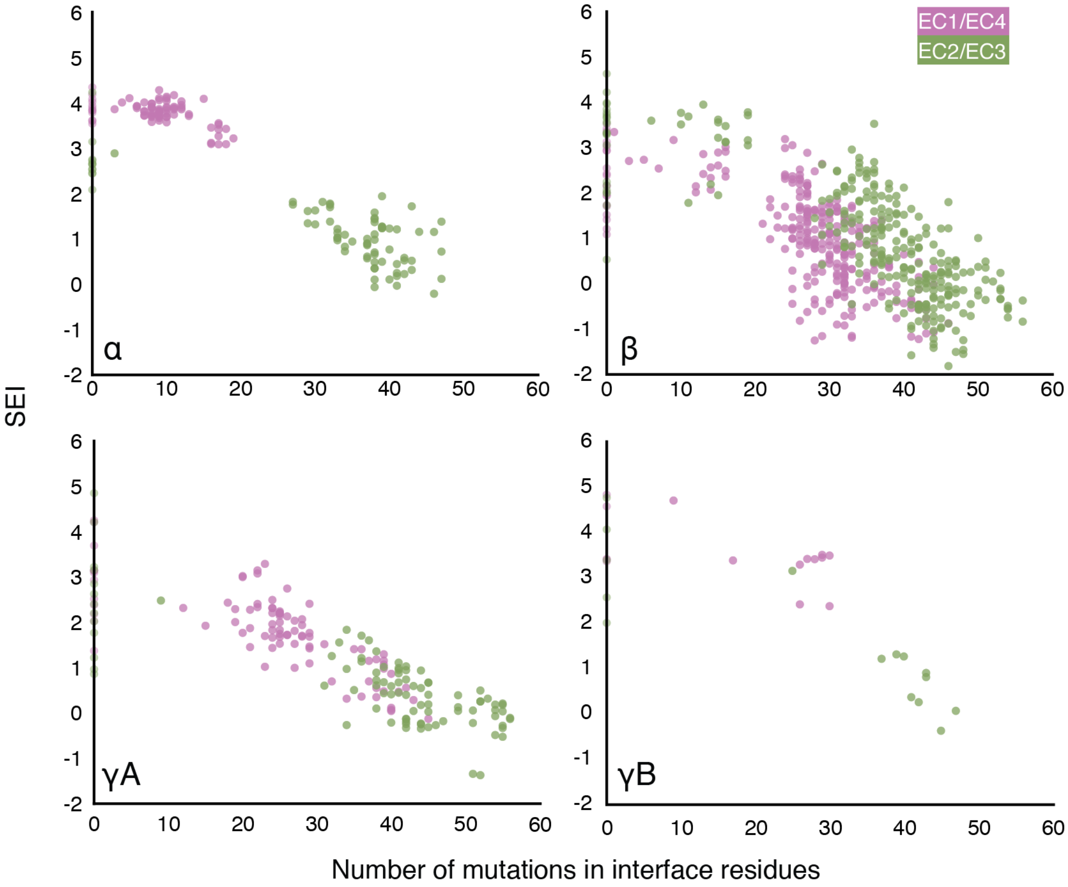


Fig. S11. Statistical energy correlates with number of amino acid substitutions
The four plots represent each possible self and nonself interaction of clustered Pcdhs within a clustered Pcdh family. The *x*-axis shows the number of mutations that separate each nonself pair, and the *y*-axis shows the statistical energy of interaction (SEI) of the corresponding interaction interface. In general, the more mutations separate the two Pcdh pairs, the lower the predicted statistical energy.

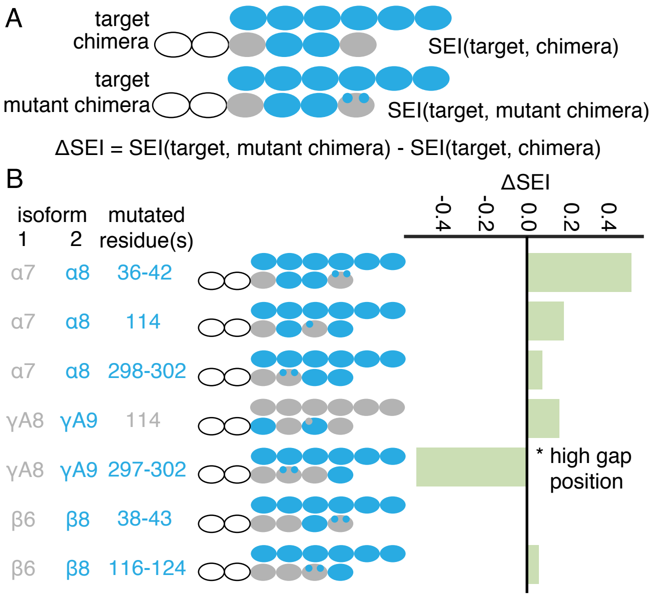


Fig. S12. Statistical energy of interaction in chimera mutants
(A) Schematic of calculation of statistical energy difference between chimera sequences and mutant chimera sequences. (B) Chimeras from (1) and our computed statistical energy for each pairing. Mutated residues are numbered according to Figure 5 of (1). The γA8/γA9 297-302 pair was not predicted well due to gaps in the alignment in this residue range.

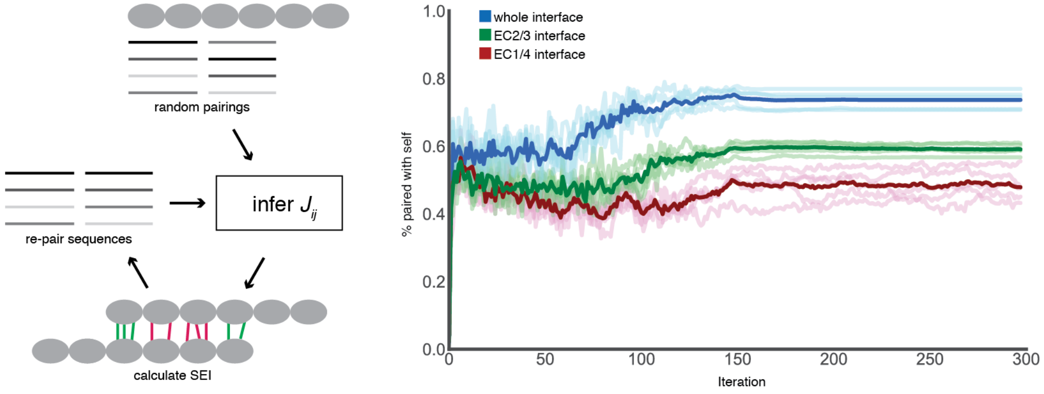


Fig. S13. Iterative pairing algorithm

(A) Schematic of the iterative pairing algorithm. After randomly pairing starting sequences, the parameters of the evolutionary couplings model are inferred and used to update the pairings. Each cycle of inference and update is one iteration. (B) Results of five replicates where the evolutionary couplings from the whole interface, just the EC2/EC3 interface, or just the EC1/EC4 interface are used to update the alignments. Light lines indicate individual runs, dark runs are the mean of the replicates.

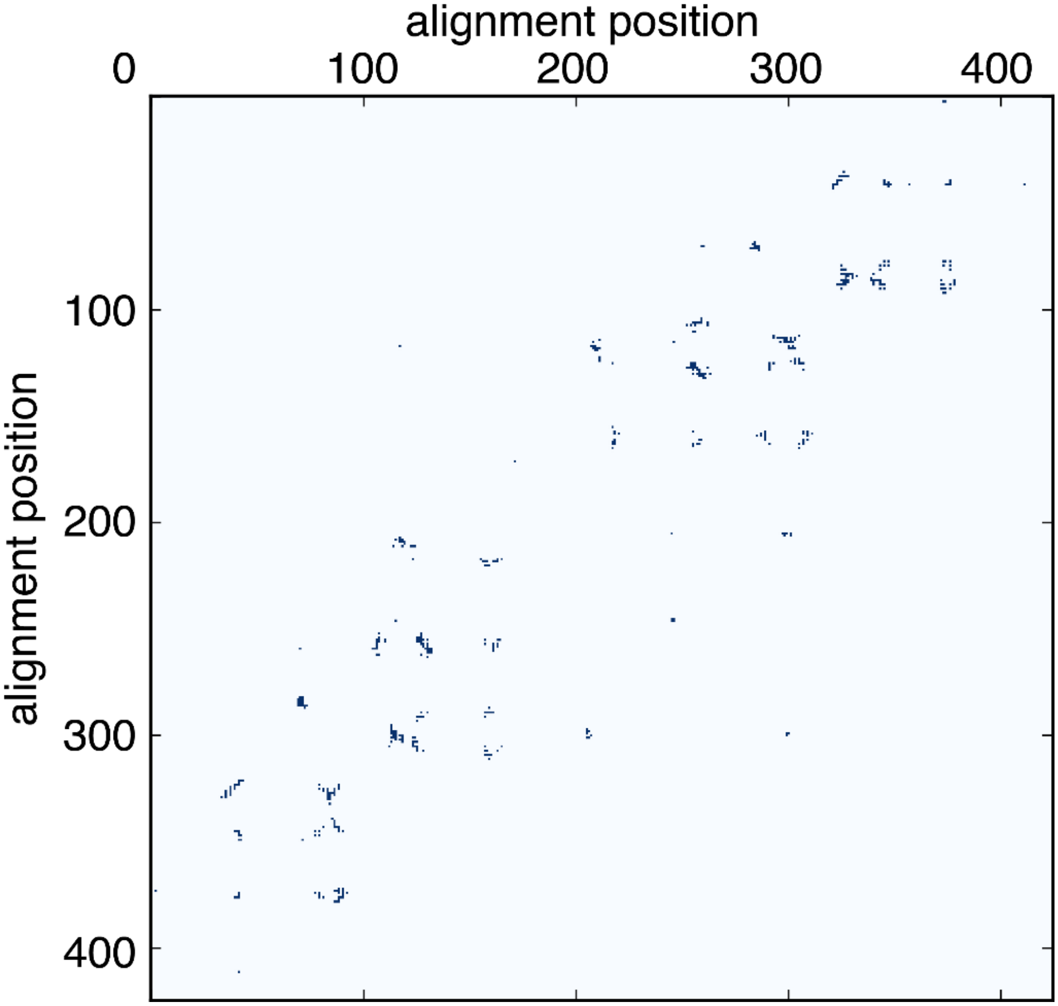


Fig. S14. Interface contact map from Pcdh simulations

Dimer interface contact map, where each dot represents two residues that come within 5 Å of each other (distance measured at the closest pair of non-hydrogen atoms) for at least 10% of the simulations of human PcdhγB3 EC1-4, mouse PcdhγB7 EC1-4, mouse Pcdhα7 EC1-5, or mouse Pcdhβ6 EC1-4. The resulting 1880 pairs arise from a total of 236 residues which comprise the interface residues used to generate the statistical interaction energy model.

Table S1. Overview of MD simulations

Label	PDB	t_{sim} (ns)	Size (# atoms)	Size (nm^3)
Mouse PCDHα7 EC1- EC5	5dzv	120	500,917	52.4 × 9.94 × 9.45
Mouse PCDHβ6 EC1-EC4	5dzx	120	394,722	41.5 × 9.86 × 9.48
Human PCDHγB3 EC1-EC4	5k8r	120	335,584	48.9 × 8.21 × 8.21
Mouse PCDHγB7 EC1-EC4	5szp	120	308,093	43.5 × 8.84 × 7.88

Table S2. BSA of individual EC interactions in MD simulations

Isoform	Interacting pair	BSA (Å^2)	Overall BSA (Å^2)
PcdhγB3	EC1/EC4	800 ± 200	4400 ± 400
	EC2/EC3	1200 ± 200	
	EC3/EC2	1100 ± 200	
	EC4/EC1	900 ± 100	
Pcdhα7	EC1/EC4	700 ± 200	3500 ± 200
	EC2/EC3	900 ± 100	
	EC3/EC2	600 ± 100	
	EC4/EC1	1000 ± 300	
Pcdhβ6	EC1/EC4	1000 ± 100	4200 ± 300
	EC2/EC3	1000 ± 100	
	EC3/EC2	1100 ± 100	
	EC4/EC1	900 ± 100	
PcdhγB7	EC1/EC4	1100 ± 100	4500 ± 300
	EC2/EC3	1100 ± 200	
	EC3/EC2	900 ± 200	
	EC4/EC1	900 ± 100	

Table S3. Data statistics for low HEPES and HEPES-free PcdhγB3 EC1-4 structures

Protein	PcdhγB3 EC1-4 no HEPES	PcdhγB3 EC1-4 less HEPES
PDB ID	6meq	6mer
SBGridDB ID	602	603
Data Collection		
Beam source	APS 24-ID-C	APS 24-ID-C
Wavelength (Å)	1.07	0.98
Space group	C222$_1$	C222$_1$
Unit cell (a, b, c; Å)	128.39, 161.77, 52.16	126.81, 162.91, 52.86
Unit cell (α, β, γ)	90, 90, 90	90, 90, 90
Resolution (Å)	28.61-3.0 (3.107-3.0)	46.74-2.9 (3.004-2.9)
Total reflections	39257 (3576)	61580 (2636)
Unique reflections	10828 (1046)	11833 (819)
Multiplicity	3.6 (3.4)	5.2 (3.2)
Completeness (%)	95.75 (93.97)	94.21 (67.19)
Mean I/σ(I)	9.36 (1.60)	10.02 (1.45)
Wilson B-factor	82.12	75.67
R$_{merge}$	0.147 (0.987)	0.113 (0.680)
R$_{meas}$	0.167 (1.15)	0.125 (0.793)
CC$_{1/2}$	0.988 (0.51)	0.996 (0.617)
CC*	0.997 (0.822)	0.999 (0.874)
Refinement		
Refinement resolution range	28.61-3.0 (3.107-3.0)	46.74-2.9 (3.004-2.9)
Reflections used in refinement	10825 (1045)	11831 (819)
Reflections used for R-free	1089 (104)	1181 (78)
R$_{work}$	0.225 (0.336)	0.224 (0.377)
R$_{free}$	0.272 (0.390)	0.271 (0.446)
CC$_{work}$	0.952 (0.638)	0.940 (0.600)
CC$_{free}$	0.911 (0.449)	0.877 (0.279)
Number of non-hydrogen atoms	3326	3330
Macromolecules	3219	3229
Ligands (Ca^{2+})	9	9
Waters	98	92
Protein residues	414	416
RMS Bonds (Å)	0.003	0.003
RMS Angles (°)	0.90	0.88
Clashscore	3.45	3.12
Average B-factor	92.38	92.28
Macromolecules	93.47	93.25
Ligands	74.16	75.77
Solvent	58.32	59.78
Ramachandran plot regions		
Favored (%)	96.60	96.38
Allowed (%)	3.40	3.62
Outliers (%)	0	0
Rotamer outliers (%)	0.28	1.93

Table S4. Interface residues from Pcdh structures and simulations

Grouping	number of pairs	number of positions
crystal 5 Å*	542	141
crystal 8 Å*	1634	205
MD 10%‡	1880	236
MD 5%‡	2096	243
MD 0%‡	3042	276

* Residue pairs that are within 5 or 8 Å of each other in crystal structures

‡ Residue pairs that are within 5 Å of each other for at least 0%, 5% or 10% of the simulation frames

Additional dataset S1 (separate file)

List of interface residues from molecular dynamics simulations based on PcdhγB3 numbering

Additional dataset S2 (separate file)

List of interface residue pairs used in statistical energy of interaction model based on PcdhγB3 numbering

Additional dataset S3 (separate file)

Alignments of mouse clustered Pcdh isoforms

References

1. Rubinstein R, et al. (2015) Molecular Logic of Neuronal Self-Recognition through Protocadherin Domain Interactions. *Cell* 163(3):629–642.
2. Chen W V., et al. (2012) Functional Significance of Isoform Diversification in the Protocadherin Gamma Gene Cluster. *Neuron* 75(3):402–409.
3. Li Y, et al. (2012) Molecular and Functional Interaction between Protocadherin- C5 and GABAA Receptors. *J Neurosci* 32(34):11780–11797.
4. Mah KM, Houston DW, Weiner JA (2016) The γ-Protocadherin-C3 isoform inhibits canonical Wnt signalling by binding to and stabilizing Axin1 at the membrane. *Sci Rep* 6(1):31665.
5. Nicoludis JM, et al. (2016) Antiparallel protocadherin homodimers use distinct affinity- and specificity-mediating regions in cadherin repeats 1-4. *Elife* 5:e18449.
6. Otwinowski Z, Minor W (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology, Volume 276: Macromolecular Crystallography, Part A*, eds Carter Jr. CW, Sweet RM (Academic Press, New York), pp 307–326. 276th Ed.
7. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* 60(12):2126–2132.
8. Adams PD, et al. (2010) PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 66(2):213–221.
9. Morin A, et al. (2013) Collaboration gets the most out of software. *Elife* 2013(2):1–6.
10. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802.
11. Huang J, MacKerell AD (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J Comput Chem* 34(25):2135–2145.
12. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual Molecular Dynamics. *J Mol Graph* 14:33–38.
13. Nicoludis JM, et al. (2015) Structure and Sequence Analyses of Clustered Protocadherins Reveal Antiparallel Interactions that Mediate Homophilic Specificity. *Structure* 23:2087–2098.
14. Goodman KM, et al. (2016) Structural Basis of Diverse Homophilic Recognition by Clustered α- and β-Protocadherins. *Neuron* 90(4):709–723.
15. Goodman KM, et al. (2016) γ-Protocadherin structural diversity and functional implications. *Elife* 5:e20930.
16. Johnson LS, Eddy SR, Portugaly E (2010) Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* 11(1):431.
17. Bitbol A-F, Dwyer RS, Colwell LJ, Wingreen NS (2016) Inferring interaction partners from protein sequences. *Proc Natl Acad Sci* 113(43):12180–12185.
18. Gueudré T, Baldassi C, Zamparo M, Weigt M, Pagnani A (2016) Simultaneous identification of specifically interacting paralogs and interprotein contacts by direct coupling analysis. *Proc Natl Acad Sci* 113(43):12186–12191.
19. Morcos F, et al. (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci* 108(49):E1293–E1301.
20. Marks DS, et al. (2011) Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLoS One* 6(12):e28766.
21. Hopf TA, et al. (2018) The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics* (October):326918.