Title: "Structural origins of clustered protocadherin-mediated neuronal barcoding"

Authors: Rotem Rubinstein[1,2†], Kerry Marie Goodman[1†], Tom Maniatis[1,3*], Lawrence Shapiro[1,2,3*], and Barry Honig[1,2,3,4,5*].

Affliations: [1]Department of Biochemistry and Molecular Biophysics, [2]Department of Systems Biology, [3]Zuckerman Mind Brain and Behavior Institute, [4]Howard Hughes Medical Institute, and [5]Department of Medicine, Columbia University, New York, NY 10032, USA.

[†] Co-first authors

[*] Correspondence: tm2472@cumc.columbia.edu (T.M.); shapiro@convex.hhmi.columbia.edu (L.S.); bh6@cumc.columbia.edu (B.H.)

Tel: 212-851-4652

Postal address: Columbia University, 1130 St. Nicholas Avenue, Room 815, New York, NY 10032

Abstract (100 words)

Clustered protocadherins mediate neuronal self-recognition and non-self discrimination—neuronal "barcoding"—which underpin neuronal self-avoidance in vertebrate neurons. Recent structural, biophysical, computational, and cell-based studies on protocadherin structure and function have led to a compelling molecular model for the barcoding mechanism. Protocadherin isoforms assemble into promiscuous *cis*-dimeric recognition units and mediate cell-cell recognition through homophilic *trans*-interactions. Each recognition unit is composed of two arms extending from the membrane proximal EC6 domains. A *cis*-dimeric recognition unit with each arm coding adhesive *trans* homophilic specificity can generate a zipper-like assembly that in turn suggests a chain termination mechanism for self-vs-non-self-discrimination among vertebrate neurons.

Keywords (5)

Clustered protocadherins, Neuronal self-avoidance, Cell-cell recognition, Protein interaction specificity, Crystal structure

1. Introduction

The establishment of functional neural circuits in the human brain involves highly specific connections among billions of neurons through trillions of synapses [1]. The formation of such complex neural circuits depends on a limited repertoire of guidance cues and cell surface receptors. Clustered protocadherins (Pcdhs) are a family of highly diverse cell-surface receptors that are thought to provide individual neurons with single-cell-specific molecular "barcodes" that provide unique cell surface identities required for neuronal self-avoidance [2-4]. Although recent publications have demonstrated that Pcdhs have additional roles— in regulating neuronal survival, synaptogenesis, dendritic arborization, and neuronal tiling [5-15, 2, 3, 16, 17, 4] —this review focuses primarily on the role of Pcdhs in neuronal self-avoidance that in turn requires that neurons be able to distinguish "self" from "non-self".

Mammalian genomes contain 50–60 Pcdhs that are arranged in three contiguous gene clusters designated α, β, and γ [18, 19]. Each Pcdh isoform has a distinct extracellular region, single pass transmembrane helix, and short cytoplasmic region encoded by a single variable exon. Additionally, the α- and γ-clusters each contain three constant exons that encode a cluster-specific constant cytoplasmic region. Phylogenetic analysis of the 58 clustered Pcdh mouse isoforms revealed that they fall into five distinct subfamilies (Figure 1): alternate α-Pcdhs (1–12), alternate β-Pcdhs (1–22), alternate γA-Pcdhs (1–12), alternate γB-Pcdhs (1–2 & 4–8), and C-type Pcdhs (αC1, αC2, γC3, γC4, and γC5) (Figure 1). Alternate (non-C-type) Pcdh isoforms are chosen for expression in each neuron by a stochastic promoter choice mechanism [19-24]. Individual neurons appear to express a

small subset of the ~50 alternate isoforms [19-24]. The C-type Pcdhs appear to be subject to a distinct mechanism of gene expression regulation and are expressed 'deterministically' rather than stochastically [22, 23].

In neuronal self-avoidance, an essential feature of neural circuit assembly, branching neurites (axons and dendrites) from the same neuron avoid one another, while neurites from different neurons do not. This assures that neurites from the same neuron can arborize extensively while neurites from different neurons can interdigitate and occupy the same field. This phenomenon requires a mechanism to allow individual neurons to distinguish self from non-self interactions [25, 26]. It appears that, for both vertebrates and insects, neuronal self-avoidance relies on generating unique individual cell surface identities through the stochastic expression of diverse repertoires of cell surface protein isoforms [27, 28, 25, 26]. In the fly, it has been demonstrated that neuronal identity is defined by the expression of single-cell-specific subsets of Dscam1 isoforms, generated by stochastic alternative RNA splicing [29-32]. In vertebrates, it appears that neuronal identity is provided by stochastic expression of single-cell-specific subsets of Pcdh isoforms [22-24, 4].

Counter-intuitively, in both insects and vertebrates the process of self-avoidance begins with adhesive homophilic interactions required for recognition [33-35, 26, 36]. In the fly, there are 19 008 possible Dscam isoforms with distinct extracellular domains, of which ~10–50 are expressed in each neuron [37, 29, 33, 30, 31]. The majority of these isoforms bind in *trans* in a strictly homophilic manner [33, 34]. In vertebrates the 50–60 Pcdh isoforms have been shown to bind with homophilic specificity, as will be discussed below. Current thinking posits that identical Dscam/Pcdh isoforms located on the surface of neurites emanating from the same cell will bind to each other homophilically in *trans* (different neurites) and will trigger a signaling process, which requires the intracellular domains [38], which leads to repulsion. In contrast, when two neurons expressing a sufficiently diverse set of Dscam/Pcdh isoforms come into contact, their different

isoform composition will not lend itself to homophilic binding and hence an avoidance mechanism will not be triggered [26].

The large number of potential Dscam1 isoforms decreases the probability that any two interacting fly neurons will have an identical or even a similar isoform repertoire [28]. Assuming, for example, that 15 distinct isoforms are chosen per cell, the probabilities that two cells will express three or more isoforms in common (thereby presumably leading to inappropriate initial adhesion and then repulsion) is ~10^{-7} (Table 1). These numbers are small enough to ensure that inappropriate repulsion will be a rare event [28]. How do Pcdhs, with far fewer isoforms, provide sufficient diversity for a single-cell identity within mammalian nervous systems which are far more complex than that of the fly? Recent structural and biophysical studies combined with cell assays have provided a surprising mechanism to answer this question. Here, we review these studies that have transformed our understanding of Pcdh structure and function and that have led to development of a structure-based mechanism for neuronal barcoding that allows Pcdhs to achieve even greater neuronal diversity than *Drosophila* Dscam1.

2. Homophilic cell-cell recognition specificity

In common with many other cadherin superfamily members [39], Pcdhs function in cell-cell recognition through binding between their extracellular regions in a manner that is calcium dependent [36]. The Pcdh extracellular region contains six extracellular cadherin (EC) domains, each of which is composed of approximately 100 residues forming a two-layered anti-parallel β-sheet structure. Binding three Ca^{2+} ions to cadherin-conserved calcium-binding motifs stabilizes Pcdh EC interdomain junctions.

In an important study, using quantitative cell aggregation assay with K562 cells, Schreiner & Weiner [35] tested seven γ-Pcdhs and showed that they exhibit isoform-specific homophilic binding. Schreiner and Weiner [35] showed that K562 cells were a suitable cell line for protocadherin expression, as they are non-adherent in culture and do not endogenously express protocadherins or other cell adhesion molecules. In these assays plasmids expressing individual Pcdh isoforms are transfected into the K562 cells and *trans*-binding is assayed by cell aggregation (Figure 2). Thu *et al.* [36] extended these studies and showed that Pcdh isoforms from all three gene clusters mediate specific homophilic interactions. In these studies mixing two cell populations transfected with identical isoforms results in mixed aggregates. In contrast, mixing two cell populations transfected with different isoforms resulted in separate homophilic aggregates (Figure 2). Remarkably, even when the transfected isoforms had greater than 90% sequence identity there was no observed cross binding [36].

Notably, Pcdh γC4 and all α-Pcdhs fail to reach the cell surface when expressed alone [40, 41, 36] and therefore cannot mediate cell aggregation [36]. However, when these isoforms are co-transfected with any β, γ or some C-type Pcdh isoforms (carrier Pcdhs) they are able to reach the cell surface and are thus able to mediate cell aggregation [36]. In fact, co-transfection of γC4 or any α-Pcdh with fragments that include the EC5–EC6 domains of carrier Pcdhs is sufficient for cell-surface delivery [36]. Importantly, the observation that γC4 and α-Pcdhs are carried to the cell surface by other Pcdhs indicates a *cis* (same cell) interaction between carrier Pcdhs and γC4 and α-Pcdhs that is dependent on EC5–EC6 domains (this will be further elaborated below) [36]. Overall, with the exception of Pcdh-αC1, all mouse Pcdh isoforms bind homophilically between apposed cells [35, 36].

3. Crystal structures of Pcdh *trans* dimers

The most thoroughly characterized cadherins are the classical cadherins, which mediate calcium dependent cell-cell adhesion through *trans* (cell-cell) homodimerization of their membrane-distal EC1 domains. In contrast to classical cadherins the first Pcdh structures obtained, which included the membrane-distal EC1 domains [42] or EC1–EC3 domain fragments [43, 44], were found to be monomeric in solution. Consistently, constructs of corresponding size did not mediate cell-cell binding in cell aggregation assays [44]. These early Pcdh structures revealed that despite containing cadherin domains, Pcdhs are structurally distinct from classical cadherins. Most notably, the first β-strand (A-strand) of EC1 lacks the critical Trp-2 residue, which is conserved among classical and desmosomal cadherins and anchors the strand-swap *trans*-binding interface of these cadherins [45, 42-44, 46]. In addition, the inter-domain orientation of the three EC domains within each structure results in an overall straight architecture. This is in contrast to the curved architecture of classical cadherins which facilitates the formation of a parallel EC1/EC1 interaction forming between molecules from opposed cell surfaces [45, 43, 44].

Rubinstein *et al.* [44] demonstrated through solution biophysical measurements and cell aggregation assays with a Pcdh-ectodomain truncation series that EC1–EC4 was required for Pcdh *trans*-binding. They then used docking calculations of the EC1–EC3 structures, constrained by sequence and mutagenesis experiments to determine that the EC1–EC4 domains form an extended *trans*-binding interface that is topologically similar among all clustered Pcdhs. Moreover, the analysis also strongly suggested that the *trans*-interaction occurs in a head-to-tail (anti-parallel) arrangement, with EC1 interacting with EC4 and EC2 interacting with EC3. Independently, Nicoludis *et al.* [43] used Pcdh EC1–EC3 fragment structures which they had determined along with correlated mutation analysis to arrive at a similar conclusion.

More recent papers have presented crystal structures of Pcdh EC1–EC4-mediated *trans*-dimers describing the *trans* adhesive interface in atomic detail. Crystal structures of ectodomain fragments corresponding to EC domains 1–4 or 1–5 for nine different isoforms, including at least two representative isoforms from each of the four Pcdh subfamilies (α, β, γA, and γB), have been solved [47-49] (Figure 3A). Despite significant diversity in their sequences, isoforms from all subfamilies formed structurally similar dimers (Figure 3A). For almost all structures, membrane-distal EC domains 1–4 dimerize in a head-to-tail orientation in which residues from EC1 domains contact residues from EC4 domains and residues from EC2 domains contact residues from EC3 domains (Figure 3A). Individual molecules from each complex were found to be highly structurally similar to the monomeric EC1–EC3 structures, indicating that complex formation did not involve any significant structural rearrangements [47]. The structures revealed that the interface is not continuous but is instead divided between an EC1/EC4 interface and an EC2/EC3 interface (Figure 3B and 3C). The average buried surface area upon *trans* dimer formation was found to be 4666 $Å^2$ with the interface between EC1/EC4 and EC2/EC3 burying on average 2062 and 2604 $Å^2$ respectively [47, 48].

Despite having a similar domain structure, the homodimeric antiparallel EC1–EC4 interface formed by the clustered Pcdhs is fundamentally different from the homodimeric interfaces of classical cadherins and the heterodimeric interfaces of cadherin-23/Pcdh-15 and desmosomal cadherins (Figure 4) [47-49]. The classical and desmosomal cadherins bind by swapping their N-terminal β-strand (A-strand) between two interacting EC1 domains (Figure 4) [45, 50]. In addition, an "X-dimer" interface, located in the linker region between EC1 and EC2, functions as a binding intermediate in classical cadherin and is the actual adhesive interface in the classical cadherin related protein T-cadherin (Figure 4) [51, 52]. Similar to the Pcdhs, the hetero-dimeric complex between cadherin-23 and Pcdh-15 exhibits an antiparallel interface; however, this interface comprises the EC1 and EC2 domains and hence is distinct from the clustered Pcdh interfaces

(Figure 4) [53]. Recently, the crystal structure of the homodimeric complex of Pcdh-19, a non-clustered δ2-Pcdh, revealed an anti-parallel EC1–EC4 interface that is highly similar to that of the clustered Pcdh *trans*-dimer, thereby indicating that in addition to clustered Pcdhs and Pcdh-19 this interface may be used by other non-clustered δ-Pcdhs [54]. Overall, the cadherin domain has demonstrated a remarkable diversity in its binding mechanisms.

4. Structural basis of Pcdh homophilic specificity

In order to identify the Pcdh *trans*-homophilic specificity-determining domains Pcdh chimeras with shuffled EC domains between different isoforms were used [35, 44]. Studies using Pcdh chimeras with multiple domains shuffled simultaneously demonstrated that chimeras with non-matching EC1 and EC4 domains do not bind to each other even when their EC2 and EC3 domains are identical [44]. By contrast, when chimeras have containing all matching four EC1 through EC4 domains identical they do, in fact, bind to each other. Together these and other data from mutagenesis experiments showed that all four membrane distal domains EC1–EC4 contribute to binding specificity [44, 47, 48].

The atomic-resolution structures of Pcdh *trans*-dimers of representative isoforms from α, β, and γ clusters accompanied by bioinformatics analyses have yielded significant insights into how Pcdhs achieve their remarkable *trans*-homophilic specificity. The structures of Pcdh *trans*-dimer complexes are similar overall among all isoforms, most notably in the antiparallel interaction between EC1–EC4 domains (Figure 3A). However, isoforms from different clusters generally exhibit prominent local structural differences in their dimerization interfaces. These are likely the primary reason why α/β, α/γA, β/γA, and β/γB heterodimers do not form [47, 48].

By contrast, the homodimeric structures of isoforms from the same cluster are structurally similar both globally and locally [47, 48]. The basis of specific homophilic binding preferences within subfamilies is therefore not predominantly architectural, but rather due to differences in the interfacial residues. In the mouse, over 90% of the interface residues, across all four interfacial domains, exhibit sequence variation among isoforms of the β and γ cluster (Figure 3C) [47, 48]. Importantly, many of these residues that vary among the mouse isoforms are conserved among different species suggesting that these residues play an important role in recognition specificity [47, 48]. Interfacial residues within the EC2–EC3 domains of α isoforms also exhibit high sequence variability among mouse isoforms with a similar isoform specific conservation pattern among different species. However, the EC1 and EC4 domains of α isoforms are exceptionally conserved with over 90% of interfacial residues conserved among all 12 mouse isoforms [47]. Specifically, for the EC1 domain, only Pcdh-α8 exhibits variability in its interfacial residues compared to the other isoforms, with two interfacial residues showing Pcdh-α8-specific conservation [44, 47]. The conservation of the EC1/EC4 interface in mouse α-Pcdh isoforms is suggestive of functional role unique to α-isoforms.

Residue-swap experiments, in which interfacial residues that exhibit isoform-specific conservation were shuffled between isoforms, confirmed that such residues underpin Pcdh *trans* recognition specificity. These experiments also demonstrated that generation of new homophilic specificities often requires swapping pairs or small groups of residues that interact with one another in the *trans*-interface [44, 47, 48]. Mutated isoforms did bind homophilically but no longer bind to their wild-type parent isoforms [47]. The overall logic of generating strict homophilic specificity between closely related isoforms involves a relatively small number of interactions that are favorable in homodimers but unfavorable in heterodimers. In some cases these correspond to stabilizing salt bridges in the homodimer that would be disrupted in the putative heterodimer resulting in

electrostatic repulsion. In others shape complementarity in the homodimer is replaced by steric hindrance in the heterodimer. Overall, these results are consistent with the free energy of binding being distributed over four interfaces with the presence of all four domains necessary to generate sufficient affinity to produce a stable homodimer [44, 47, 48].

5. Interference and Tolerance

It is critical that two different neurons not recognize each other as "self". However, since both Dscams and Pcdhs are stochastically expressed, there is a finite probability that any pair of neurons will express one or more common isoforms, which will then bind to each other and potentially signal both cells to move apart. How can this inappropriate repulsion be avoided? Table 1 reports probabilities that two cells will randomly express one or more of the same isoforms for both Dscams and Pcdhs (assuming 15 different isoforms are expressed per cell). Even for *Drosophila* Dscam1, with thousands of isoforms to select from, the probability that two neurons will select at least one identical isoform is relatively high [28] while for Pcdhs, with only 58 isoforms to choose from it is essentially a certainty (Table 1). Therefore it is critical for interacting neurons to be able to tolerate the presence of some common isoforms without triggering repulsion. What is the maximum proportion of common expressed isoforms between two interacting neurons that can be tolerated (the "tolerance") before the two cells recognize each other erroneously as self? Figure 5A illustrates two extreme cases: the first is where a single isoform that is shared between two interacting neurons is sufficient to trigger repulsion even if all other isoforms are different (no tolerance to common isoforms, panel i); the second is when two interacting neurons repel each other only if all their expressed isoforms are identical and where a single isoform mismatch is sufficient to prevent erroneous repulsion (high tolerance to common isoforms, panel ii).

It is important to note that for *Drosophila* Dscam1, with thousands of isoforms, a tolerance of 20% was assumed [28]. As seen in Table 1, two interacting neurons will have a probability of only about 10^{-7} to share more than 20% (3 out of 15) of isoforms in common, which appears to be sufficiently rare to prevent inappropriate repulsion between interaction neurons in *Drosophila* [28]. In contrast, Pcdhs would have to have a tolerance of about 80% (12 of 15 isoforms) to achieve a similar probability ($\sim 10^{-7}$) of inappropriate pairwise repulsion.

What mechanism underlies tolerance? For Dscams it is not hard to imagine that a small fraction of common isoforms (e.g. 3 out of 15) is too small for two cells to adhere. For example, there may simply not be enough Dscams expressed on the cell surface to achieve a functionally useful adhesion complex. But this logic clearly fails for Pcdhs so that some other mechanism must be involved. An important clue is provided by cell aggregation assays, which revealed that cells expressing multiple Pcdh isoforms will only co-aggregate with cells expressing the identical set of isoforms [36] (Figure 5B). Remarkably, when 4 isoforms are expressed per cell even a single mismatch is sufficient to "interfere" with cell-cell aggregation (Figure 5B). Notably, N-cadherin does not interfere with Pcdh-mediated aggregation (Figure 5C) [36].

The interference seen in cell assays (Figure 5B) suggests that Pcdhs have a tolerance of at least 75%, which generates a pairwise probability of inappropriate recognition similar to that of Dscams with a tolerance of 20% (Table 1, [36]). It is possible that the tolerance is even higher, e.g. if a single mismatch would be enough to interfere with recognition when 10 isoforms are expressed per cell (tolerance of 90%). In this way, two different cells would incorrectly recognize each other as "self" only if their full complement of isoforms was identical (rather than merely "similar" as is the case for Dscam1). We now turn to a discussion of structural studies that have revealed how this level of tolerance can be achieved.

6. *Cis*-dimeric recognition units

In addition to their homophilic *trans* interactions Pcdhs also interact in *cis* [35, 36, 44, 48]. Solution biophysical measurements of purified recombinant Pcdh ectodomains and cell aggregation studies showed that Pcdhs form *cis* dimers mediated by EC5 and EC6. Specifically, ectodomain fragments containing the EC5–EC6 domains of β and γB isoforms dimerize in solution independent of the EC1–EC4 *trans* dimer interactions but do not aggregate cells indicating that this interaction occurs *in cis* [44, 48] (Figure 6A). Larger *cis*-multimers (e.g. tetramers) had been inferred in earlier studies [35, 55] however only *cis*-dimers have been conclusively observed.

The important role of EC6 for *cis*-dimer formation can be seen from the fact that while wild-type Pcdh γB6 behaves as a tetramer in solution (a *trans* dimer of *cis* dimers), a single point mutant in the EC6 domain that breaks the *cis* interaction behaves as a dimer in solution [48]. In cell aggregation assays this point-mutant prevents γB6 Pcdh from both self-delivering and delivering α-Pcdhs to the surface, thereby suggesting that *cis* dimerization is required for cell surface delivery [48]

A large body of evidence has accumulated demonstrating that Pcdh *cis*-interactions are promiscuous, with evidence for the formation of both homo- and hetero-*cis* dimers: First, co-immunoprecipitation experiments involving isoforms from different clusters showed promiscuous interactions between isoforms of the β and γB cluster, isoforms of γA and γB clusters, and isoforms of the γA cluster and the γC3 isoform [35, 36]. Second, in isolation, neither Pcdh-γC4 nor α-Pcdh isoforms can reach the cell surface; however, they are successfully delivered to the cell surface by co-expression with αC2, γC3, or any β or γ isoform (‘carrier Pcdhs’) [36]. Cell aggregation assays with either truncated isoforms or isoforms with shuffled domains have demonstrated that cell surface delivery depends on the EC5–EC6 domains of both the alpha and the carrier Pcdhs.

Importantly, the identity of the specific α-isoform or the specific carrier isoform used in the cell assays does not appear to impact the outcome, thereby indicating a promiscuous interaction between α (or γC4) Pcdhs and carrier Pcdhs [36]. Third, the sequences of the EC6 domains, which control *cis* dimerization, are highly similar among β-isoforms and among γ-isoforms [36], a finding that is consistent with the idea that promiscuous interactions occur within each family.

Notably, in spite of their general promiscuity, not all possible *cis*-dimers form and the homophilic binding affinities of *cis*-dimers from different subfamilies are highly variable. As previously mentioned α-Pcdhs and Pcdh-γC4 do not reach the cell surface when expressed singly. This is likely due to a failure of these isoforms to form *cis*-homodimers. In addition, while β-Pcdhs, γB-Pcdhs and Pcdh-αC2 form strong *cis*-homodimers (8.92 – 80.1 μM), with comparable affinities to the *trans* interaction affinity of Pcdhs and other cell adhesion molecules. γA-isoforms and Pcdh-γC3 do not form *cis*-homodimers with measureable affinities in solution [44, 48]. *Cis* interactions of classical cadherins are similarly weak in solution, however in the two-dimensional environment of the plasma membrane the Pcdh *cis* interactions will likely be further enhanced [56, 57]. Together these data suggest that *cis* associations between isoforms from different families will manifest different preferences [36, 44, 48]**.** Identifying preferences of *cis* interactions and their molecular origins is of great importance since, alongside Pcdh expression data, it will determine the repertoire of Pcdh *cis*-dimer recognition units presented on the cell surface [36, 44, 48].

7. Two models of the Pcdh recognition complex and associated functional implications

Two molecular models have been proposed to account for the role of Pcdhs in neuronal recognition. The first was originally based on the assumption that Pcdhs form *cis*-tetrameric recognition units that interact in *trans* to form discrete octamers between apposed cells [35, 55]. In this model,

interference is caused by the dilution of a matched isoforms pair on different cells through their incorporation into a large number of *cis* tetramers with isoforms that are not matched [55]. However, although the tetramer/octamer dilution model succeeds in providing a sufficient level of diversity to account for non-self discrimination [55], it fails in that there is also dilution within a single cell such that the probability of two sister neurites containing the same tetramer, and hence repelling, would be much too small [44]. Of course the model is in any case disproved by the fact that Pcdhs form *cis*-dimers, not tetramers [44, 48].

The dilution model can also be applied to dimeric recognition units which form a *trans* tetramer between apposed cells. However, *cis*-dimers do not appear to provide sufficient diversity to account for non-self recognition and indeed all models based on the existence of discrete oligomeric recognition units encounter difficulties in explaining both self recognition and non-self discrimination [44]. Thus, it appears that the notion of achieving diversity through the formation of discrete Pcdh multimeric recognition units (where each unit plays the same role as a Dscam monomer, (Figure 6B, middle) is insufficient to account for the role of Pcdhs in neuronal barcoding.

In the second model, the "isoform-mismatch chain termination" model (Figure 6B, right), each "arm" of a Pcdh *cis*-dimeric recognition unit interacts with one "arm" from *two* different recognition units on the apposed cell surface to form a one-dimensional zipper or lattice-like structure [44]. When identical isoforms are present in both cells, the length of the zipper-like chain is limited only by the copy number of the expressed isoforms. In contrast, cells with even a single mismatch will have the growing chain terminated by the incorporation of an isoform with no match on the apposing cell so that only small Pcdh assemblies will be formed (Figure 7) [44]. Statistical modeling illustrates that this chain-termination model exhibits step-function-like behavior which could yield a binary on/off signal (Figure 7) [44]. Underlying this model is the hypothesis that assembly size plays a crucial role in signaling such that large assemblies would transduce an

intracellular signal initiating repulsion, while the signal from small assemblies formed in the presence of a mismatch would remain below a critical threshold (Figure 7) [44]. The 'isoform-mismatch chain termination' model, in principle, can provide a mechanism for self-avoidance with self/non-self discrimination power higher than that achieved by the 19 008 distinct isoforms of *Drosophila* Dscam1 [44].

8. Conclusion

Comparison of the molecular logic of Dscams and Pcdhs reveals a number of remarkable insights as to how vertebrates and many invertebrates have evolved to solve the problem of neuronal barcoding. *Drosophila* use alternative splicing to generate diversity that is coded on three independent domains, each of which exhibits homophilic binding specificity. Since each domain presents a separate interface, the 19 008 distinct isoforms simply corresponds to the product of the number of alternative exons that can be expressed for each domain. Vertebrate diversity is based on stochastic promoter choice which does not lend itself to the combinatorial diversity that can be generated via alternative splicing. Consequently, vertebrates have had to evolve a very different mechanism for neuronal barcoding.

A combination of structure determination, biophysical measurements, cell aggregation studies and computational analysis has revealed a likely molecular mechanism. Pcdhs appear on the cell surface as *cis* dimers that contain two distinct arms, each with its own *trans* homophilic binding specificity. Two sister neurites from the same neuron will contain the same complement of Pcdh isoforms and hence an identical, or near identical, population of *cis* dimers. When two sister neurites come into contact the two sets of *cis* dimers will form *trans* interactions that produce a linear assembly whose dimensions depend on the total number of expressed Pcdhs. The model assumes that this large assembly then produces a signal for the two cells to move apart, with the assembly itself likely

[3] Kostadinov D, Sanes JR. Protocadherin-dependent dendritic self-avoidance regulates neural connectivity and circuit function. Elife. 2015;4.

[4] Mountoufaris G, Chen WV, Hirabayashi Y, O'Keeffe S, Chevee M, Nwakeze CL, et al. Multicluster Pcdh diversity is required for mouse olfactory neural circuit assembly. Science. 2017;356:411-4.

[5] Wang X, Weiner JA, Levi S, Craig AM, Bradley A, Sanes JR. Gamma protocadherins are required for survival of spinal interneurons. Neuron. 2002;36:843-54.

[6] Weiner JA, Wang XZ, Tapia JC, Sanes JR. Gamma protocadherins are required for synaptic development in the spinal cord. Proceedings of the National Academy of Sciences of the United States of America. 2005;102:8-14.

[7] Fukuda E, Hamada S, Hasegawa S, Katori S, Sanbo M, Miyakawa T, et al. Down-regulation of protocadherin-alpha A isoforms in mice changes contextual fear conditioning and spatial working memory. Eur J Neurosci. 2008;28:1362-76.

[8] Hasegawa S, Hamada S, Kumode Y, Esumi S, Katori S, Fukuda E, et al. The protocadherin-alpha family is involved in axonal coalescence of olfactory sensory neurons into glomeruli of the olfactory bulb in mouse. Mol Cell Neurosci. 2008;38:66-79.

[9] Lefebvre JL, Zhang Y, Meister M, Wang X, Sanes JR. gamma-Protocadherins regulate neuronal survival but are dispensable for circuit formation in retina. Development. 2008;135:4141-51.

[10] Prasad T, Wang X, Gray PA, Weiner JA. A differential developmental pattern of spinal interneuron apoptosis during synaptogenesis: insights from genetic analyses of the protocadherin-gamma gene cluster. Development. 2008;135:4153-64.

[11] Garrett AM, Weiner JA. Control of CNS synapse development by {gamma}-protocadherin-mediated astrocyte-neuron contact. J Neurosci. 2009;29:11723-31.

[12] Lin C, Meng S, Zhu T, Wang X. PDCD10/CCM3 acts downstream of {gamma}-protocadherins to regulate neuronal survival. J Biol Chem. 2010;285:41675-85.

[13] Prasad T, Weiner JA. Direct and indirect regulation of spinal cord la afferent terminal formation by the gamma-Protocadherins. Frontiers in Molecular Neuroscience. 2011;4.

[14] Chen WV, Alvarez FJ, Lefebvre JL, Friedman B, Nwakeze C, Geiman E, et al. Functional Significance of Isoform Diversification in the Protocadherin Gamma Gene Cluster. Neuron. 2012;75:402-9.

[15] Hasegawa S, Hirabayashi T, Kondo T, Inoue K, Esumi S, Okayama A, et al. Constitutively expressed Protocadherin-alpha regulates the coalescence and elimination of homotypic olfactory axons through its cytoplasmic region. Front Mol Neurosci. 2012;5:97.

[16] Hasegawa S, Kumagai M, Hagihara M, Nishimaru H, Hirano K, Kaneko R, et al. Distinct and Cooperative Functions for the Protocadherin-α,-β and-γ Clusters in Neuronal Survival and Axon Targeting. Frontiers in Molecular Neuroscience. 2016;9:155.

[17] Chen WV, Nwakeze CL, Denny CA, O'Keeffe S, Rieger MA, Mountoufaris G, et al. Pcdhalphac2 is required for axonal tiling and assembly of serotonergic circuitries in mice. Science. 2017;356:406-11.

[18] Kohmura N, Senzaki K, Hamada S, Kai N, Yasuda R, Watanabe M, et al. Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. Neuron. 1998;20:1137-51.

destroyed by proteolysis. In contrast, when neurites from different neurons come into contact, there will be a high probability of at least one mismatched isoform which is enough to limit assembly size, and no repulsion signal will be produced. Remarkably, this mechanism produces an essentially binary signal that could be used to distinguish self from non-self (Figure 7).

Of course much remains to be done to validate and refine the zipper/chain-termination model, and ultimately to test it in neurons, but it is consistent with all available data and explains how 58 mouse Pcdhs can code for greater diversity than 19 000 fly Dscams. This model explains the function of the stochastically expressed alternate Pcdhs, however the C-type Pcdhs, which engage in similar protein interactions to the alternate Pcdhs, appear to be expressed deterministically rather than stochastically suggesting they play distinct functional roles. Exploration of this will require further study. More generally, the results of the past few years illustrate the importance of a greater integration of structural biology, molecular biophysics and neurobiology as molecular structure can suggest novel mechanisms that would be hard to otherwise imagine.

9. Acknowledgments

This work was supported by a National Science Foundation grant to B.H. (MCB-1412472), an NIH grant to L.S. (R01GM062270), and a joint NIH grant to T.M. and L.S. (R01GM107571). The computing in this project was supported by two NIH instrumentation grants (S10OD012351 and S10OD021764) received by the Department of Systems Biology at Columbia University.

10. References

[1] Azevedo FA, Carvalho LR, Grinberg LT, Farfel JM, Ferretti RE, Leite RE, et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J Comp Neurol. 2009;513:532-41.

[2] Lefebvre JL, Kostadinov D, Chen WV, Maniatis T, Sanes JR. Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. Nature. 2012;488:517-21.

[19] Wu Q, Maniatis T. A striking organization of a large family of human neural cadherin-like cell adhesion genes. Cell. 1999;97:779-90.

[20] Tasic B, Nabholz CE, Baldwin KK, Kim Y, Rueckert EH, Ribich SA, et al. Promoter choice determines splice site selection in protocadherin alpha and -gamma pre-mRNA splicing. Molecular Cell. 2002;10:21-33.

[21] Wang XZ, Su H, Bradley A. Molecular mechanisms governing Pcdh-gamma gene expression: Evidence for a multiple promoter and cis-alternative splicing model. Genes & Development. 2002;16:1890-905.

[22] Esumi S, Kakazu N, Taguchi Y, Hirayama T, Sasaki A, Hirabayashi T, et al. Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. Nature Genetics. 2005;37:171-6.

[23] Kaneko R, Kato H, Kawamura Y, Esumi S, Hirayama T, Hirabayashi T, et al. Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. Journal of Biological Chemistry. 2006;281:30551-60.

[24] Hirano K, Kaneko R, Izawa T, Kawaguchi M, Kitsukawa T, Yagi T. Single-neuron diversity generated by Protocadherin-beta cluster in mouse central and peripheral nervous systems. Frontiers in Molecular Neuroscience. 2012;5.

[25] Zipursky SL, Sanes JR. Chemoaffinity Revisited: Dscams, Protocadherins, and Neural Circuit Assembly. Cell. 2010;143:343-53.

[26] Zipursky SL, Grueber WB. The Molecular Basis of Self-Avoidance. In: Hyman SE, editor. Annual Review of Neuroscience, Vol 362013. p. 547-68.

[27] Hattori D, Millard SS, Wojtowicz WM, Zipursky SL. Dscam-Mediated Cell Recognition Regulates Neural Circuit Formation. Annual Review of Cell and Developmental Biology2008. p. 597-620.

[28] Hattori D, Chen Y, Matthews BJ, Salwinski L, Sabatti C, Grueber WB, et al. Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. Nature. 2009;461:644-U87.

[29] Neves G, Zucker J, Daly M, Chess A. Stochastic yet biased expression of multiple Dscam splice variants by individual cells. Nature Genetics. 2004;36:240-6.

[30] Zhan XL, Clemens JC, Neves G, Hattori D, Flanagan JJ, Hummel T, et al. Analysis of Dscam diversity in regulating axon guidance in Drosophila mushroom bodies. Neuron. 2004;43:673-86.

[31] Miura SK, Martins A, Zhang KX, Graveley BR, Zipursky SL. Probabilistic Splicing of Dscam1 Establishes Identity at the Level of Single Neurons. Cell. 2013;155:1166-77.

[32] Sun W, You X, Gogol-Doring A, He H, Kise Y, Sohn M, et al. Ultra-deep profiling of alternatively spliced Drosophila Dscam isoforms by circularization-assisted multi-segment sequencing. EMBO J. 2013;32:2029-38.

[33] Wojtowicz WM, Flanagan JJ, Millard SS, Zipursky SL. Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. Cell. 2004;118:619-33.

[34] Wojtowicz WM, Wu W, Andre I, Qian B, Baker D, Zipursky SL. A vast repertoire of Dscam binding Specificities arises from modular interactions of variable ig domains. Cell. 2007;130:1134-45.

[35] Schreiner D, Weiner JA. Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. Proceedings of the National Academy of Sciences of the United States of America. 2010;107:14893-8.

[36] Thu CA, Chen WV, Rubinstein R, Chevee M, Wolcott HN, Felsovalyi KO, et al. Single-Cell Identity Generated by Combinatorial Homophilic Interactions between alpha, beta, and gamma Protocadherins. Cell. 2014;158:1045-59.

[37] Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, et al. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell. 2000;101:671-84.

[38] Matthews BJ, Kim ME, Flanagan JJ, Hattori D, Clemens JC, Zipursky SL, et al. Dendrite self-avoidance is controlled by Dscam. Cell. 2007;129:593-604.

[39] Hulpiau P, van Roy F. Molecular evolution of the cadherin superfamily. Int J Biochem Cell Biol. 2009;41:349-69.

[40] Murata Y, Hamada S, Morishita H, Mutoh T, Yagi T. Interaction with protocadherin-gamma regulates the cell surface expression of protocadherin-alpha. J Biol Chem. 2004;279:49508-16.

[41] Bonn S, Seeburg PH, Schwarz MK. Combinatorial expression of alpha- and gamma-protocadherins alters their presenilin-dependent processing. Mol Cell Biol. 2007;27:4121-32.

[42] Morishita H, Umitsu M, Murata Y, Shibata N, Udaka K, Higuchi Y, et al. Structure of the cadherin-related neuronal receptor/protocadherin-alpha first extracellular cadherin domain reveals diversity across cadherin families. Journal of Biological Chemistry. 2006;281:33650-63.

[43] Nicoludis JM, Lau S-Y, Schaerfe CPI, Marks DS, Weihofen WA, Gaudet R. Structure and Sequence Analyses of Clustered Protocadherins Reveal Antiparallel Interactions that Mediate Homophilic Specificity. Structure. 2015;23:2087-98.

[44] Rubinstein R, Thu CA, Goodman KM, Wolcott HN, Bahna F, Mannepalli S, et al. Molecular Logic of Neuronal Self-Recognition through Protocadherin Domain Interactions. Cell. 2015;163:629-42.

[45] Boggon TJ, Murray J, Chappuis-Flament S, Wong E, Gumbiner BM, Shapiro L. C-cadherin ectodomain structure and implications for cell adhesion mechanisms. Science. 2002;296:1308-13.

[46] Harrison OJ, Brasch J, Lasso G, Katsamba PS, Ahlsen G, Honig B, et al. Structural basis of adhesive binding by desmocollins and desmogleins. Proceedings of the National Academy of Sciences of the United States of America. 2016;113:7160-5.

[47] Goodman KM, Rubinstein R, Thu CA, Bahna F, Mannepalli S, Ahlsen G, et al. Structural Basis of Diverse Homophilic Recognition by Clustered alpha- and beta-Protocadherins. Neuron. 2016;90:709-23.

[48] Goodman KM, Rubinstein R, Thu CA, Mannepalli S, Bahna F, Ahlsén G, et al. γ-Protocadherin structural diversity and functional implications. eLife. 2016;5:e20930.

[49] Nicoludis JM, Vogt BE, Green AG, Scharfe CPI, Marks DS, Gaudet R. Antiparallel protocadherin homodimers use distinct affinity- and specificity-mediating regions in cadherin repeats 1-4. eLife. 2016;5.

[50] Brasch J, Harrison OJ, Honig B, Shapiro L. Thinking outside the cell: how cadherins drive adhesion. Trends Cell Biol. 2012;22:299-310.

[51] Ciatto C, Bahna F, Zampieri N, VanSteenhouse HC, Katsamba PS, Ahlsen G, et al. T-cadherin structures reveal a novel adhesive binding mechanism. Nature Structural & Molecular Biology. 2010;17:339-U110.

[52] Harrison OJ, Bahna F, Katsamba PS, Jin X, Brasch J, Vendome J, et al. Two-step adhesive binding by classical cadherins. Nat Struct Mol Biol. 2010;17:348-57.

[53] Sotomayor M, Weihofen WA, Gaudet R, Corey DP. Structure of a force-conveying cadherin bond essential for inner-ear mechanotransduction. Nature. 2012;492:128-+.

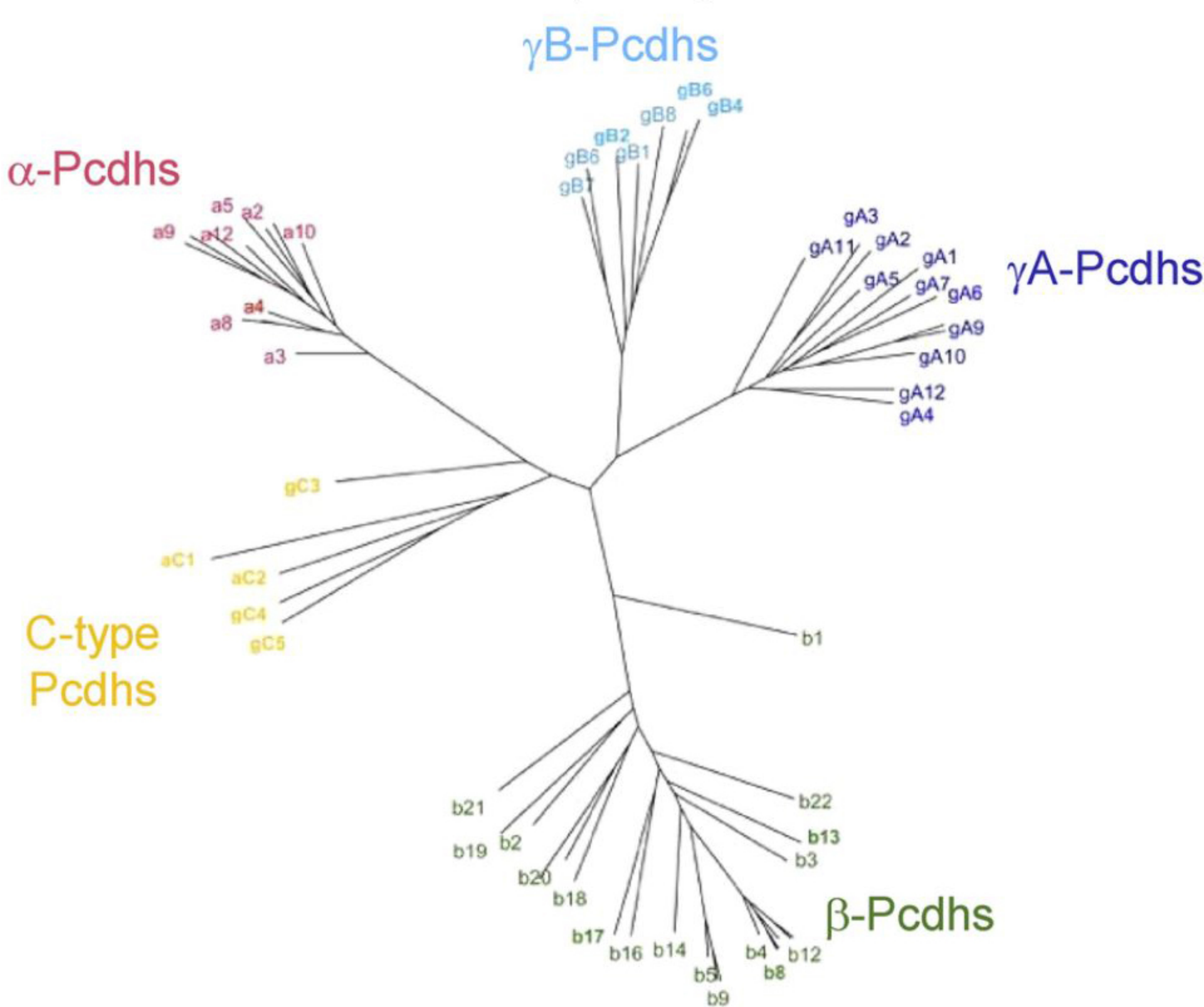
[54] Cooper SR, Jontes JD, Sotomayor M. Structural determinants of adhesion by Protocadherin-19 and implications for its role in epilepsy. Elife. 2016;5.

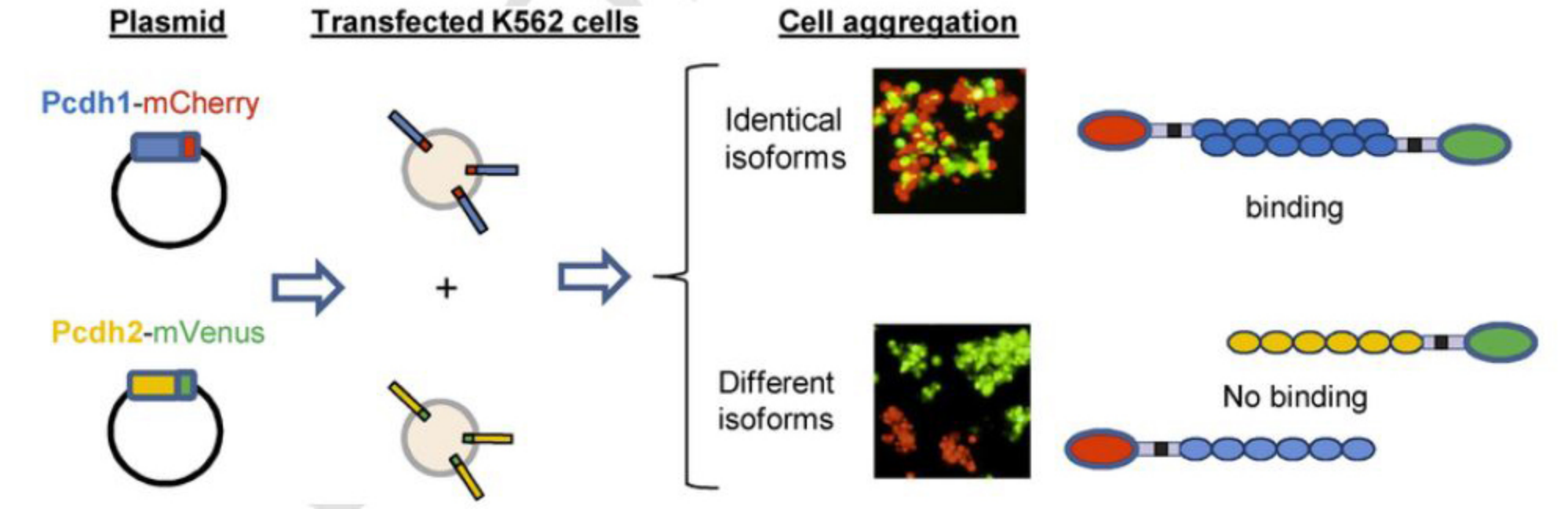
[55] Yagi T. Molecular codes for neuronal individuality and cell assembly in the brain. Front Mol Neurosci. 2012;5:45.

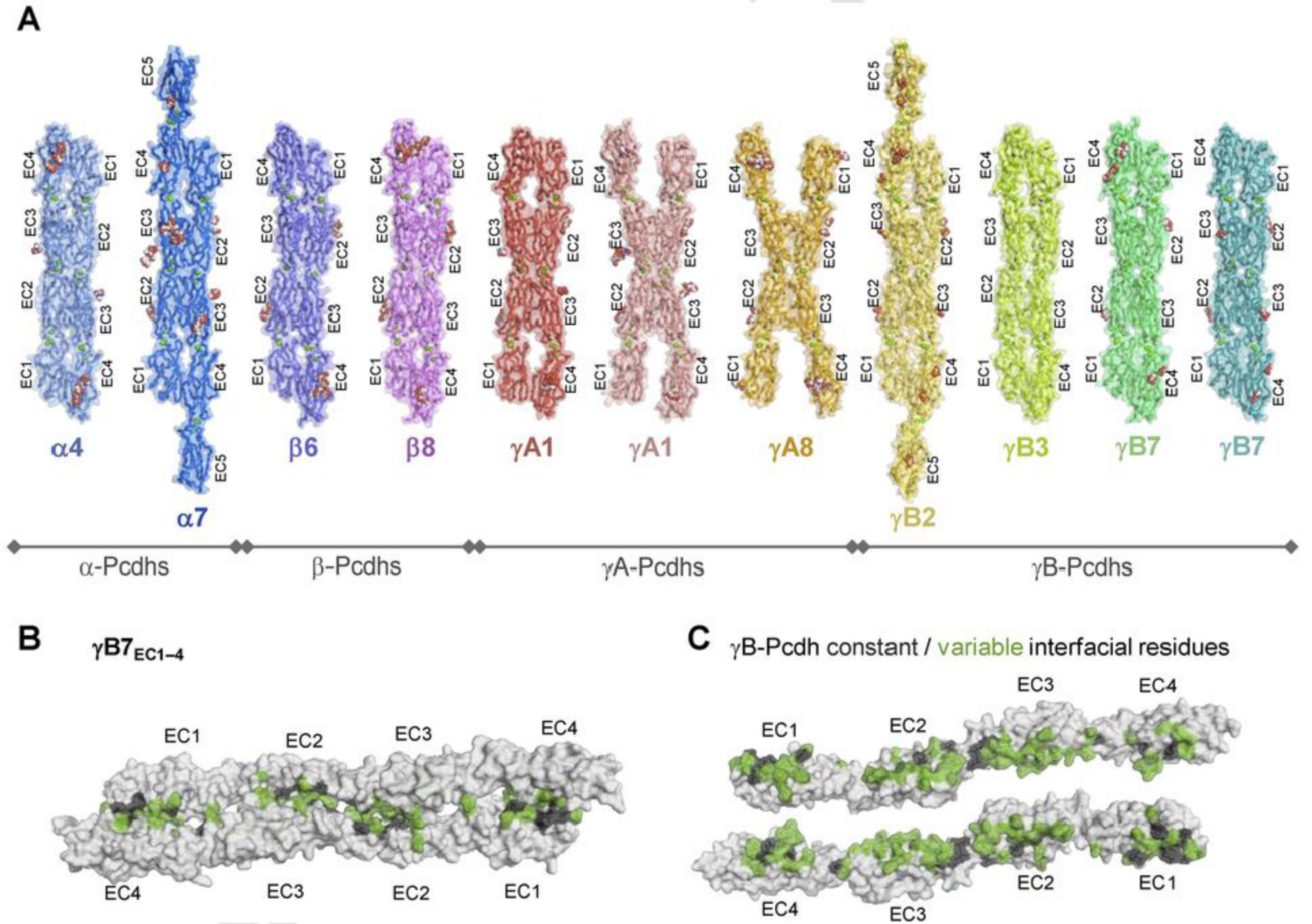
[56] Wu Y, Vendome J, Shapiro L, Ben-Shaul A, Honig B. Transforming binding affinities from three dimensions to two with application to cadherin clustering. Nature. 2011;475:510-U107.

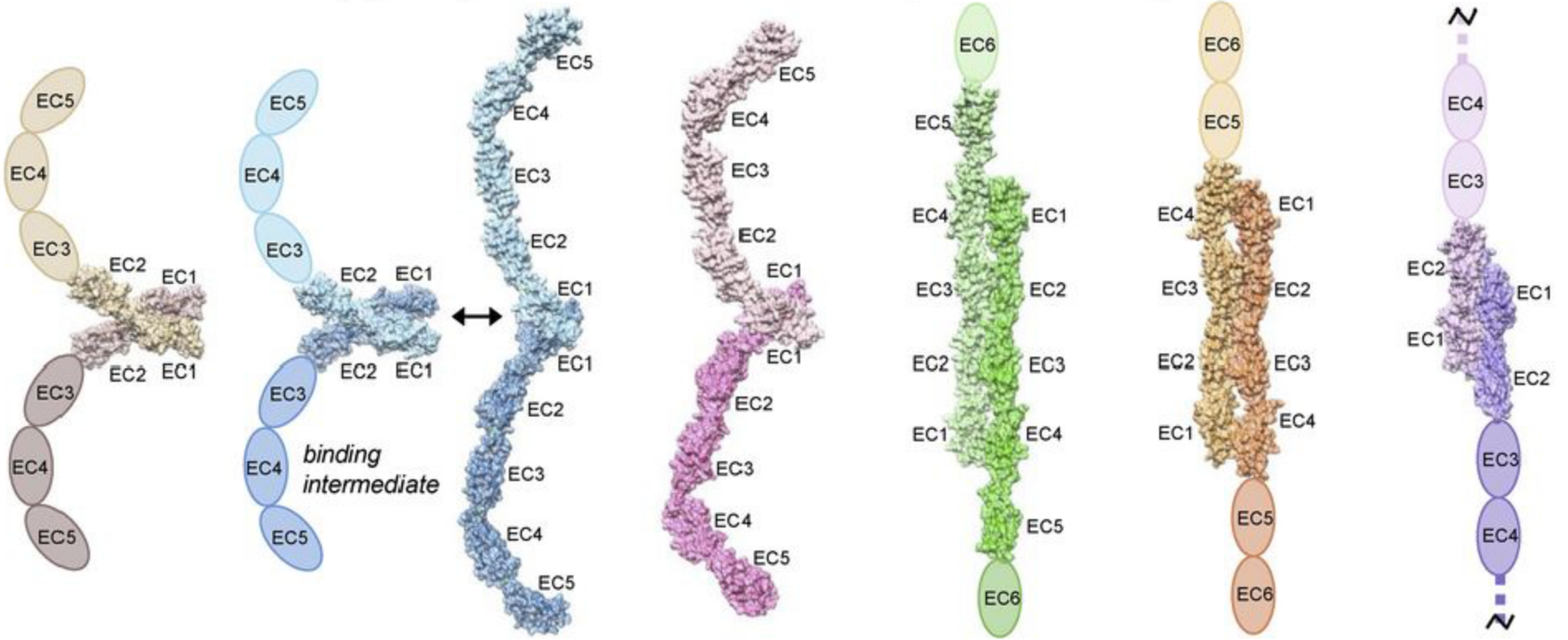
[57] Wu Y, Honig B, Ben-Shaul A. Theory and Simulations of Adhesion Receptor Dimerization on Membrane Surfaces. Biophysical Journal. 2013;104:1221-9.

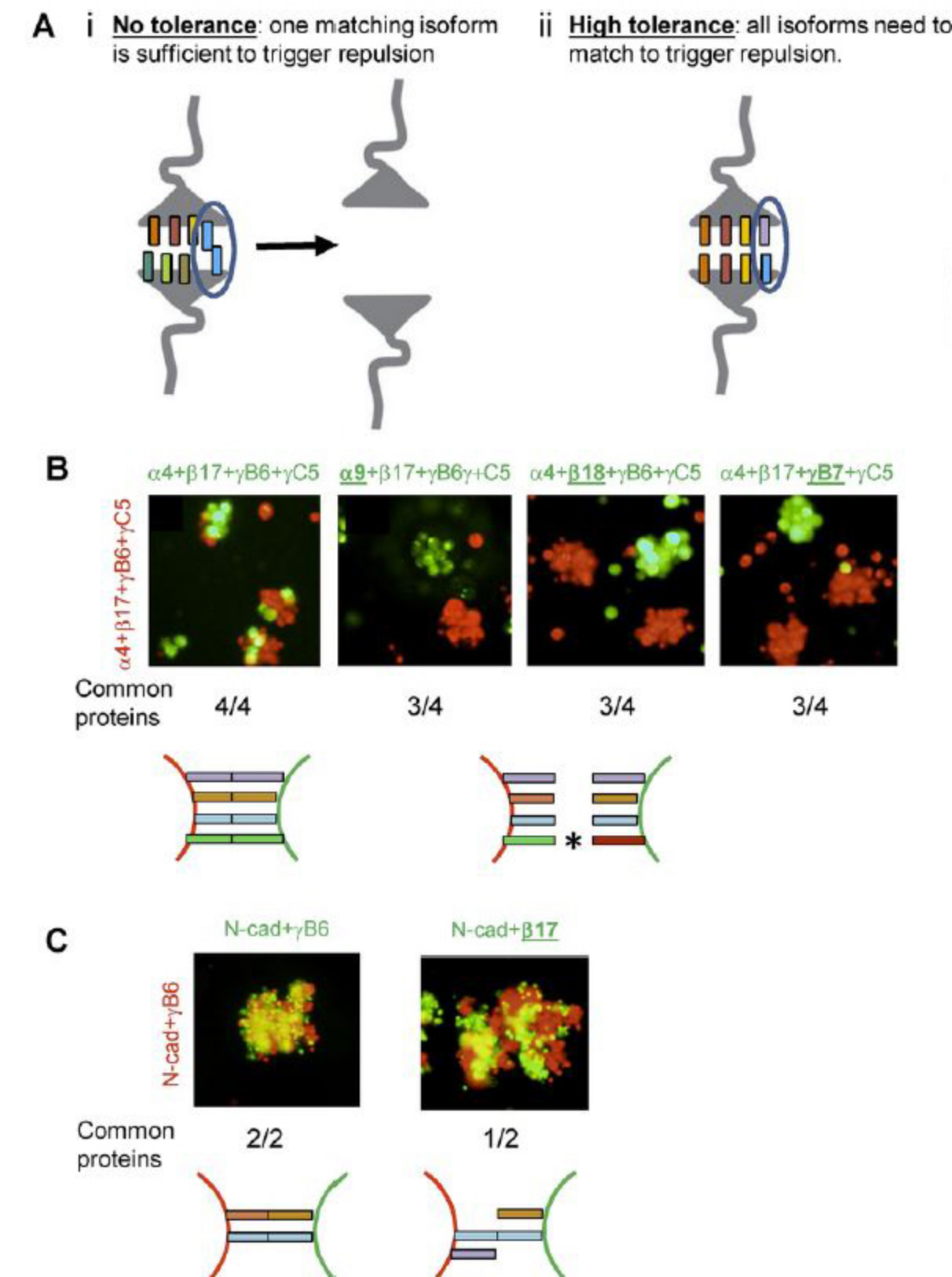
	Probability that a pair of neurons share at least N isoforms	
N, Number of common isoforms (out of 15)	Dscam	Pcdhs
1	1.18E-02	9.95E-01
2	6.07E-05	9.55E-01
3	1.80E-07	8.26E-01
4	3.41E-10	5.92E-01
5	4.34E-13	3.28E-01
6	3.81E-16	1.34E-01
7	2.32E-19	3.95E-02
8	9.79E-23	8.10E-03
9	2.81E-26	1.13E-03
10	5.32E-30	1.03E-04
11	6.37E-34	5.85E-06
12	4.47E-38	1.92E-07
13	1.63E-42	3.21E-09
14	2.45E-47	2.17E-11
15	8.61E-53	3.36E-14

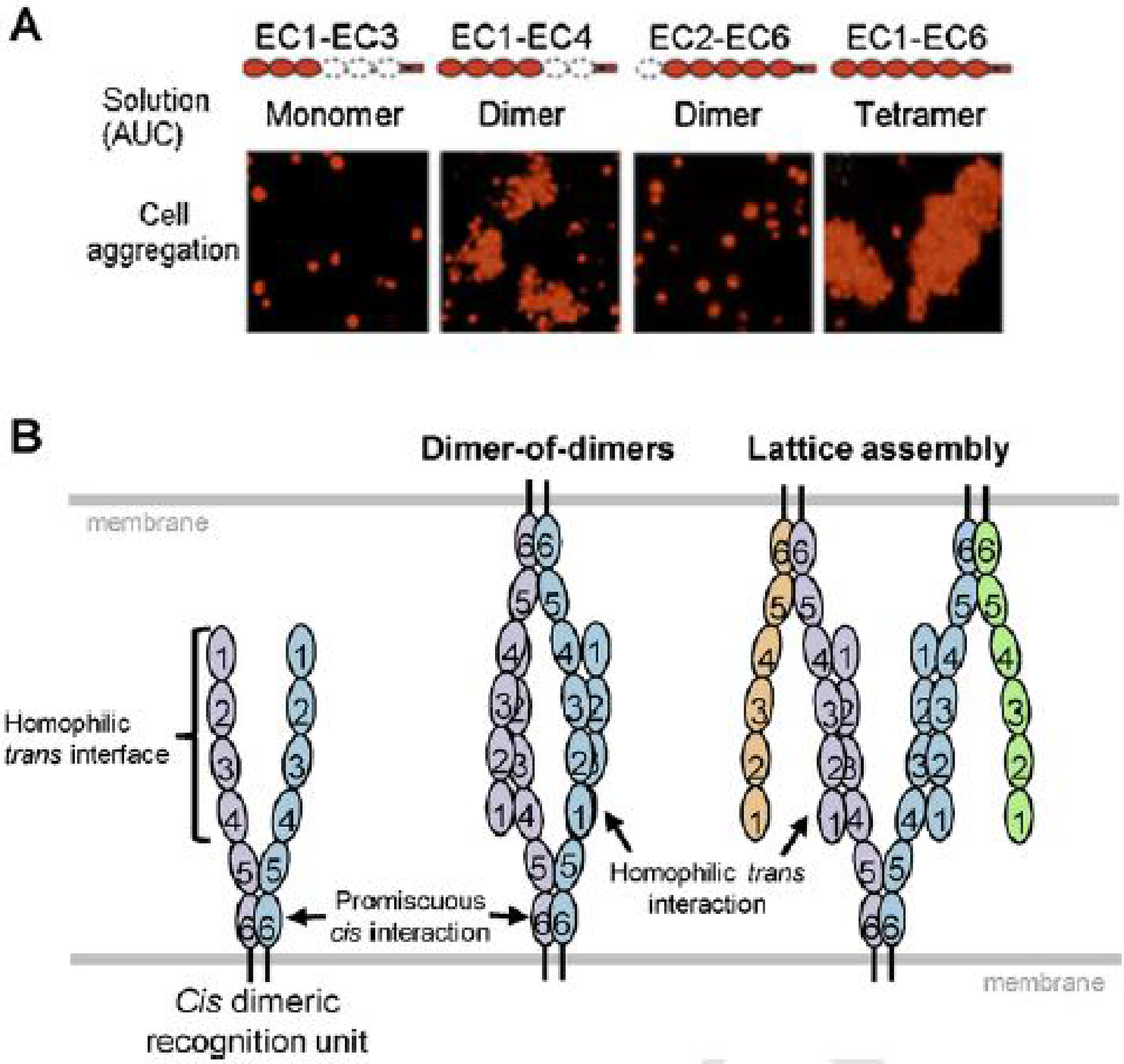

γB-Pcdhs
α-Pcdhs
γA-Pcdhs
C-type Pcdhs
β-Pcdhs
gB6
gB4
gB8
gB2
gB6
gB1
gB1
a5
a2
a9
a12
a10
a4
a8
a3
gA3
gA11
gA2
gA1
gA5
gA7
gA6
gA9
gA10
gA12
gA4
gC3
aC1
aC2
gC4
gC5
b1
b21
b22
b19
b2
b13
b3
b20
b18
b17
b16
b14
b4
b12
b5
b8
b9

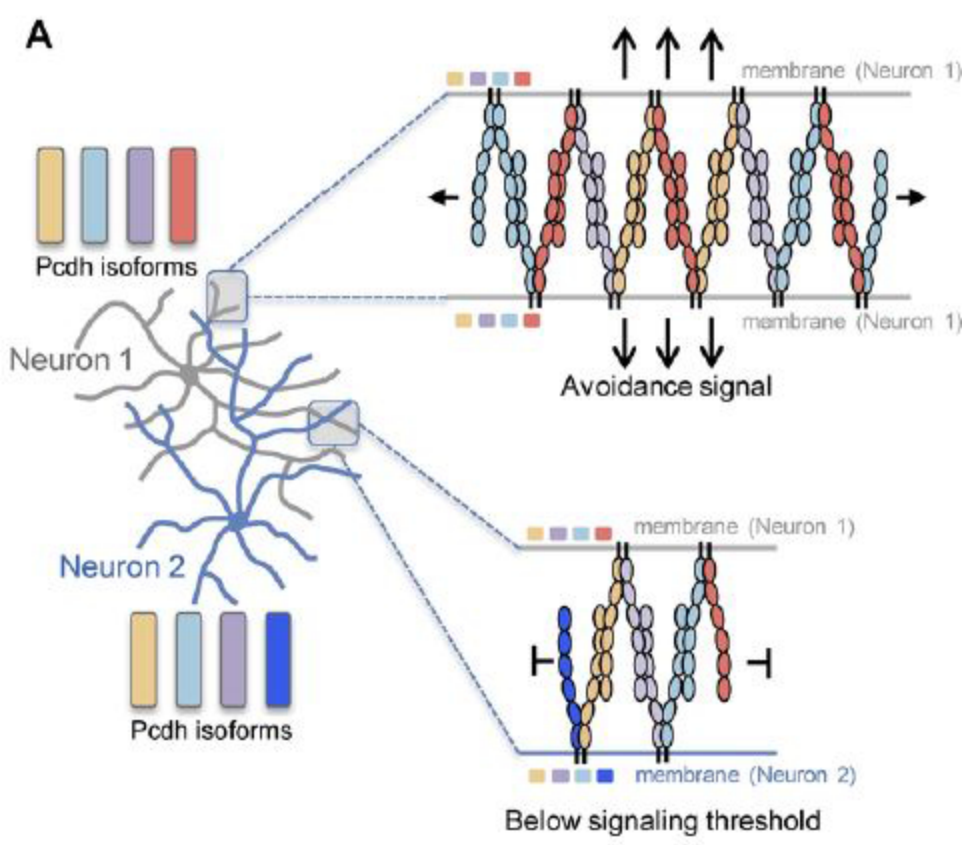

Plasmid
Transfected K562 cells
Cell aggregation
Pcdh1-mCherry
Pcdh2-mVenus
+
Identical isoforms
Different isoforms
binding
No binding

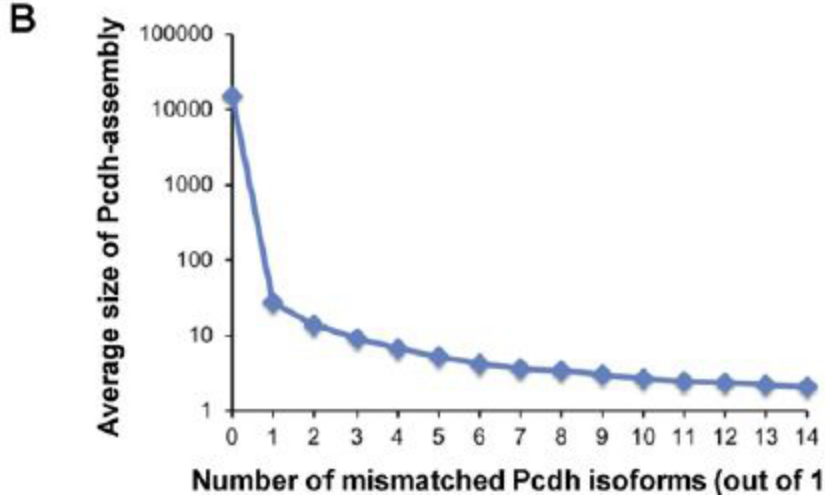

A
α4
α7
β6
β8
γA1
γA1
γA8
γB2
γB3
γB7
γB7
α-Pcdhs
β-Pcdhs
γA-Pcdhs
γB-Pcdhs
B
γB7EC1–4
EC1
EC2
EC3
EC4
C
γB-Pcdh constant / variable interfacial residues


T-cadherin
Classical cadherins
(Type I & II)
Desmosomal
cadherins
Clustered
protocadherins
Non-clustered
protocadherins
Mechanosensitive
cadherins
EC1
EC2
EC3
EC4
EC5
EC6
binding
intermediate


A
i No tolerance: one matching isoform is sufficient to trigger repulsion
ii High tolerance: all isoforms need to match to trigger repulsion.
B
α4+β17+γB6+γC5
α9+β17+γB6γ+C5
α4+β18+γB6+γC5
α4+β17+γB7+γC5
α4+β17+γB6+γC5
Common proteins
4/4
3/4
3/4
3/4
C
N-cad+γB6
N-cad+β17
N-cad+γB6
Common proteins
2/2
1/2


A
EC1-EC3
EC1-EC4
EC2-EC6
EC1-EC6
Solution (AUC)
Monomer
Dimer
Dimer
Tetramer
Cell aggregation
B
Dimer-of-dimers
Lattice assembly
membrane
Homophilic trans interface
Promiscuous cis interaction
Homophilic trans interaction
Cis dimeric recognition unit
membrane


A
membrane (Neuron 1)
Pcdh isoforms
Neuron 1
membrane (Neuron 1)
Avoidance signal
membrane (Neuron 1)
Neuron 2
Pcdh isoforms
membrane (Neuron 2)
Below signaling threshold


B
Average size of Pcdh-assembly
100000
10000
1000
100
10
1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Number of mismatched Pcdh isoforms (out of 15)