Seminars in Cell & Developmental Biology 69 (2017) 140–150

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Structural origins of clustered protocadherin-mediated neuronal barcoding

Rotem Rubinstein[a,b,1], Kerry Marie Goodman[a,1], Tom Maniatis[a,c,*], Lawrence Shapiro[a,c,*], Barry Honig[a,b,c,d,e,*]

[a] Department of Biochemistry and Molecular Biophysics, New York, NY 10032, USA
[b] Department of Systems Biology, New York, NY 10032, USA
[c] Zuckerman Mind Brain and Behavior Institute, New York, NY 10032, USA
[d] Howard Hughes Medical Institute, New York, NY 10032, USA
[e] Department of Medicine, Columbia University, New York, NY 10032, USA

ARTICLE INFO

Article history:
Received 5 June 2017
Received in revised form 14 July 2017
Accepted 18 July 2017
Available online 22 July 2017

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

ABSTRACT

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.

Contents

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", which provide unique cell surface identities required for neurite self-avoidance [2–4]. Although recent studies have demonstrated that Pcdhs have additional roles in regulating

Abbreviations: *Pcdhs*, clustered protocadherins; *EC*, extracellular cadherin domain.

* Corresponding authors at: Department of Biochemistry and Molecular Biophysics, New York, NY 10032, USA.
E-mail addresses: tm2472@cumc.columbia.edu (T. Maniatis), shapiro@convex.hhmi.columbia.edu (L. Shapiro), bh6@cumc.columbia.edu (B. Honig).
[1] Co-first authors.

http://dx.doi.org/10.1016/j.semcdb.2017.07.023
1084-9521/

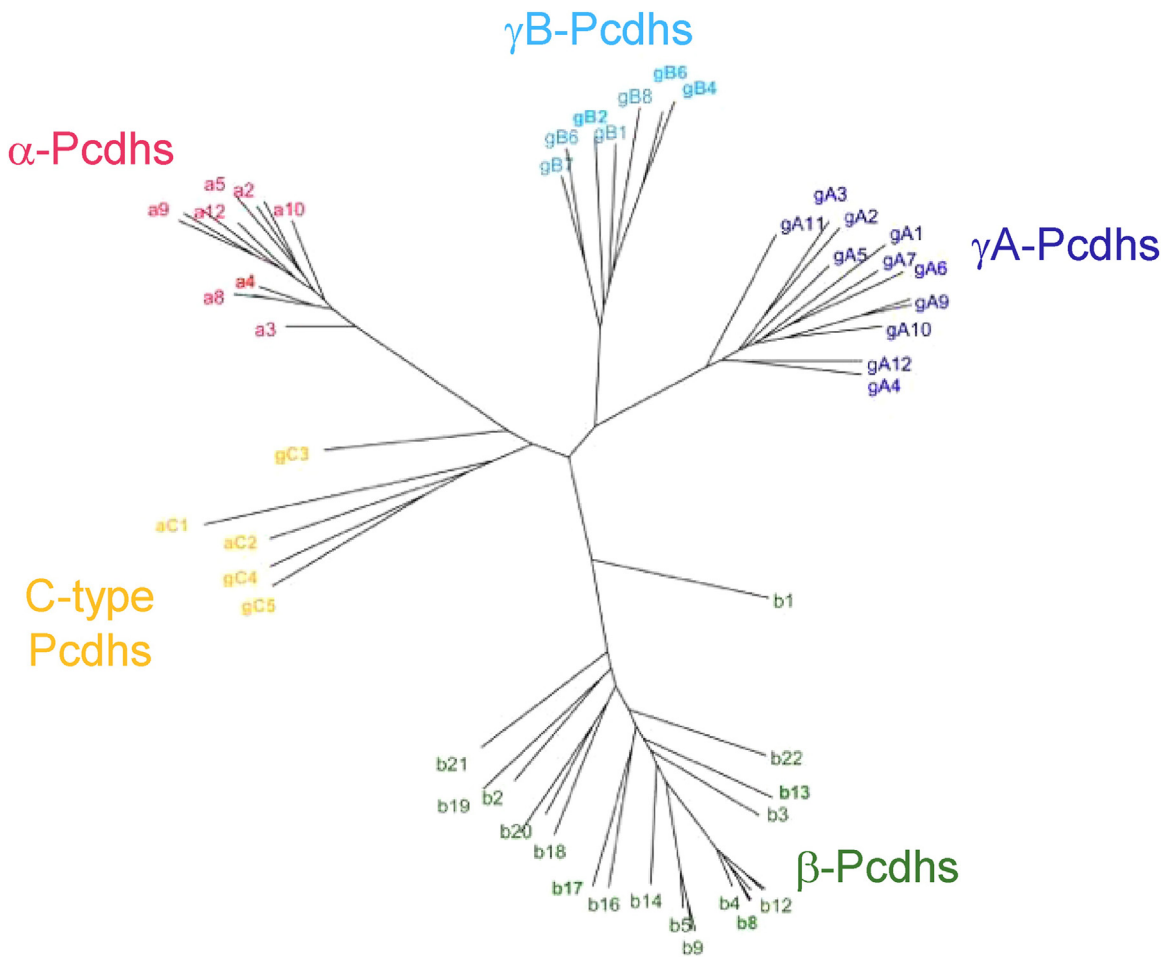

Fig. 1. Phylogenetic tree of mouse Pcdh isoforms based on the sequences of their EC1–EC4 domains which mediate the homophilic *trans*-interactions. The α, β, γA, and γB isoforms are grouped into four separate clusters. Although part of the β-cluster, the β1 isoform is significantly divergent from other members of the β-cluster. The two αC and the three γC isoforms are divergent from all other isoforms, including the other members of the α and γ clusters.

neuronal survival, synaptogenesis, dendritic arborization, and neuronal tiling [2–17] —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 *Pcdh* genes that are arranged in three contiguous gene clusters designated α, β, and γ [18,19]. Other vertebrates, such as the fugu and elephant shark, also have Pcdh genes but with varying numbers of isoforms and distinct cluster organizations [20,21]. Each Pcdh isoform has a distinct extracellular region, single pass transmembrane helix, and short cytoplasmic region encoded by a single "variable" exon. Additionally, the Pcdh α- and γ- gene 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 (Fig. 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) (Fig. 1). Alternate (non-C-type) Pcdh isoforms are chosen for expression in each neuron by a stochastic promoter choice mechanism [19,22–26]. Individual neurons appear to express a small subset of the ~50 alternate isoforms [19,22–26]. The C-type Pcdhs are expressed 'deterministically' rather than stochastically [22,24].

In neurite 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 without crossing or clumping, while neurites from different neurons can interdigitate and occupy the same field. This phenomenon requires a mechanism that allows individual neurons to distinguish self from non-self interactions [27,28]. 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 [27–30]. In the fly neuronal identity is defined by the expression of single-cell-specific subsets of Dscam1 protein isoforms, generated by stochastic alternative RNA splicing [31–34]. In vertebrates, neuronal identity is provided by stochastic expression of single-cell-specific subsets of Pcdh isoforms [4,22–24].

Counter-intuitively, in both insects and vertebrates the process of self-avoidance begins with adhesive homophilic interactions required for recognition [27,35–38]. In the fly, there are 19,008 possible Dscam isoforms with distinct extracellular domains, of which ~10–50 are expressed in each neuron [31,33–35,39]. The majority of these isoforms bind in *trans* in a strictly homophilic manner [35,36]. In mammals, 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 bind to each other homophilically in *trans* (different neurites) and this interaction triggers a signaling process that requires the intracellular domains [40], which leads to repulsion. In contrast, when two neurons expressing a sufficiently diverse set of Dscam/Pcdh isoforms