Title
Unique Identity at the Level of Single Neurons Supports Self-Recognition During the Assembly of Neural Circuits

Permalink
https://escholarship.org/uc/item/9s680071

Author
Miura, Satoru

Publication Date
2013

Peer reviewed|Thesis/dissertation
Unique Identity at the Level of Single Neurons Supports Self-Recognition During the Assembly of Neural Circuits

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Biological Chemistry

by

Satoru Miura

2013
ABSTRACT OF THE DISSERTATION

Unique Identity at the Level of Single Neurons Supports Self-Recognition During the Assembly of Neural Circuits

By

Satoru Miura

Doctor of Philosophy in Biological Chemistry
University of California, Los Angeles, 2013
Professor S. Lawrence Zipursky, Chair

From hunting preys, avoiding predators, to more complex social interactions, our behaviors depend on precise connectivity patterns of the nervous system. Nervous systems comprise numerous cell types exhibiting diverse morphologies and surprisingly stereotypical connection patterns, as Ramón y Cajal had noted more than a century ago. Its development encompasses tasks that are each enormously complex, such as cell fate determination, cell migration, axon
guidance, branch ramification, and synaptic matching. How such a feat is achieved at the molecular level continues to be a fascinating question in neurobiology. Over the last few decades, advances have been made in understanding the logic underlying how neurites are guided to their correct targets. As initially proposed by Cajal and later by Roger Sperry, the interactions of neurites with their environment are mediated by cell recognition molecules. 

*Drosophila Dscam1* encodes a set of such molecules. The gene is unique, in that it is able to generate a tremendous number of isoforms, each with unique binding specificity, from a single locus. Through mutually exclusive alternative splicing at multiple exon clusters, it has the potential to generate 19,008 extracellular domains, tethered to the membrane with one of two single-pass transmembrane domains. Extensive biochemistry has shown that each extracellular domain isoform exhibits isoform-specific homophilic binding, with little to no heterophilic binding. Genetic studies show that these isoforms mediate self-avoidance, the tendency of neurites from the same neuron to repel each other. They also suggest that neighboring neurons express distinct sets of Dscam1 isoforms, allowing the neurites to discriminate self from non-self. How is this achieved? And more importantly, could unique Dscam1 identities contribute to the assembly of neural circuits by mediating specific interactions between different neurons? To answer these questions, we developed splicing reporters to visualize the splicing pattern of one of the alternatively spliced exon clusters *in vivo*. Our
results demonstrate that the splicing in neurons identified by their unique anatomical locations is different in different animals and that it does not occur in a cell-type specific fashion. The data argue that neurons acquire unique Dscam1 identity through probabilistic splicing. In fact, probabilistic splicing was observed throughout the nervous system, suggesting that Dscam1 isoforms are unlikely to specify connectivity between different neurons. Rather, the major role of Dscam1 in the assembly of the neural circuits is to mediate self-recognition.
The dissertation of Satoru Miura is approved.

Douglas L. Black

Siavash K. Kurdistani

Kelsey C. Martin

S. Lawrence Zipursky, Committee Chair

University of California, Los Angeles

2013
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>x</td>
</tr>
<tr>
<td>Vita</td>
<td>xiii</td>
</tr>
<tr>
<td>Chapter 1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>The Roles of Cell Recognition Molecules in Wiring of the Nervous System</td>
<td>2</td>
</tr>
<tr>
<td><em>Drosophila</em> <em>Dscam1</em> Encodes a Vast Number of Isoforms That Exhibit Isoform-Specific Homophilic Binding</td>
<td>8</td>
</tr>
<tr>
<td>Multiple <em>Dscam1</em> Isoforms Are Expressed in a Single Neuron</td>
<td>12</td>
</tr>
<tr>
<td><em>Dscam1</em> Promotes Self-Avoidance in Neurons</td>
<td>11</td>
</tr>
<tr>
<td>Thousands of <em>Dscam1</em> Isoforms Are Required for Discrimination Between Self and Non-Self</td>
<td>19</td>
</tr>
<tr>
<td>Analysis of <em>Dscam1</em> Splicing: Deterministic or Probabilistic?</td>
<td>20</td>
</tr>
<tr>
<td>References</td>
<td>22</td>
</tr>
<tr>
<td>Chapter 2 Probabilistic Splicing of <em>Dscam1</em> Establishes Identity at the Level of Single Neurons</td>
<td>25</td>
</tr>
<tr>
<td>Abstract</td>
<td>26</td>
</tr>
<tr>
<td>Introduction</td>
<td>27</td>
</tr>
<tr>
<td>Results</td>
<td>31</td>
</tr>
<tr>
<td>Discussion</td>
<td>51</td>
</tr>
</tbody>
</table>
Experimental Procedures 57
Supplementary Information 62
Appendix 2-1 71
References 75

Chapter 3 Discussion 79

Probabilistic Splicing of Dscam1 Is a Robust Means to Achieve Unique Identity at the Level of Single Neurons 80
Unresolved Issues on Dscam1 Splicing 83
Evolution of Self-Avoidance and the Responsible Genes 86
References 89
LIST OF FIGURES

Chapter 1

| Figure 1-1 | Various Strategies in Neural Circuit Assembly | 5 |
| Figure 1-2 | Mutually Exclusive Alternative Splicing of *Dscam1* Generates a Vast Number of Isoforms That Exhibit Isoform-Specific Homophilic Binding | 10 |
| Figure 1-3 | Dscam1 Mediates Self-Avoidance | 15 |

Chapter 2

| Figure 2-1 | The Design of Reporters for Splicing of Alternative Exon 4 Variants | 32 |
| Figure 2-2 | Alternative Variants of Exon 4 Are Expressed at Different Frequencies in the MB | 35 |
| Figure 2-3 | Splicing of Exon 4 in Class IV da Neurons Is Probabilistic | 39 |
| Figure 2-4 | Expression of Exon 4 Variants Is Randomly Distributed Among the 42 Class IV da Neurons | 42 |
| Figure 2-5 | Splicing of Exon 4 in Class IV da Neurons Is Dynamic | 45 |
| Figure 2-6 | Scattered Neurons of Specific Cell Types Express Alternative Variants of Exon 4 in the Visual System | 48 |
| Figure 2-S.1 | Generation of Splicing Reporter Knock-Ins by Homologous Recombination Followed by phiC31 Recombination Mediated Cassette Exchange | 62 |
| Figure 2-S.2 | Expression of Alternative Variants of Exon 4 Are Variable in the MB | 63 |
Figure 2-S.3 Each Alternative Variant of Exon 4 Is Expressed in a Subset of Class IV da Neurons in a Probabilistic Fashion 65

Figure 2-S.4 No Significant Bias in the Spatial Expression Pattern of Each Alternative Variant of Exon 4 in the Class IV da Neurons Was Detected 67

Figure 2-S.5 Scattered Subsets of Specific Cell Types Express Alternative Variants of Exon 4 in the Visual System 69

Figure 2-A.1 Strategies of Ends-Out Homologous Recombination 73
ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep gratitude to my mentor, Larry Zipursky, for his guidance through my graduate training. He has had an enormous influence on how I approach science. Over the years, we had many intense and fun discussions on my project, on science in general, and on tennis. It was undeniably these stimulating conversations that pushed us to think harder and that lead to better ideas. I had the privilege to take on an exciting project that we knew from the start was going to be difficult. When experiments failed, Larry was there to support me. When experiments succeeded, he was as thrilled as I was. His confidence in me was a great encouragement. I am also grateful for the scientific environment that he has provided me. On many occasions, I was given the opportunity to interact with renowned scientists from all over the world, from whom I learned tremendously. Thank you Larry, for making my graduate training a lot of fun.

I am grateful to all the members of my committee, Doug Black, Siavash Kurdistani, and Kelsey Martin, for their assistance and for fun discussions. I always enjoyed meeting and talking with them about my project. Doug has given me numerous suggestions as an expert on mRNA splicing, Siavash always challenged my ideas until they were absolutely solid, and I had the privilege to
rotate in Kelsey's lab and be part of a very important project. I am also grateful to Brent Graveley at the University of Connecticut, with whom we collaborated on my thesis project. The project would not have been possible without his critical suggestions on experimental designs. I would like to thank Eddy De Robertis, who treated me as an honorary member of his lab. I would also like to offer special thanks to my undergraduate advisor from the University of Tokyo, Hitoshi Sakano, whose philosophy and passion in science continue to inspire me.

I could not have achieved my PhD without the support from the members of the Zipursky lab. I am grateful to Aljoscha Nern from whom I learned all the basics of fly genetics, and to Daisuke Hattori and Woj Wojtowicz for their mentorship. I would also like to thank Kelvin Zhang, André Martins, and Wei Wu with whom I worked closely, Matt Pecot for numerous intellectual conversations, and Anthony Palacios for his IT support. I am truly appreciative of all the current and past members of the lab for their tremendous inputs on my project, for good friendships, and for fun times.

I am tremendously grateful for the constant support from my family. I would like to thank my mother, Yukiko Miura, my father, Motohiro Miura, and my sister, Yuiko Miura, for their encouragement. I am also indebted to all my friends both in and
out of science for their support and friendships. My PhD would not have been possible without all the fun I had both inside and outside the lab.

My dissertation work was supported by a fellowship from the Nakajima Foundation and Dissertation Year Fellowship from UCLA.

Chapter 2 is a version of: Miura SK, Martins A, Zhang KX, Graveley BR, and Zipursky SL. Probabilistic Splicing of Dscam1 Establishes Identity at the Level of Single Neurons. Cell, in press. The authors of this manuscript thank Yasushi Hiromi, Yuh Nung Jan, Aljoscha Nern, Barret Pfeiffer, and Gerald Rubin for reagents, members of our laboratory for discussion, and Edward De Robertis, Wesley Grueber, Daisuke Hattori, Kelsey Martin, and Alex Procik for their comments on the manuscript. We also thank Orkun Akin and Josh Trachtenberg for their help with the 2-photon microscope, Jinfei Ni for his early work in generating the reporter constructs, Wei Wu for her initial assistance in assessment of the transgenic flies, and GenetiVision and Genetic Services Inc. for their injection services. This work was supported by a fellowship from the Nakajima Foundation to S.K.M., Canadian Institutes of Health Research (CIHR) Fellowship and UCLA Computational Bioscience Initiative Fellowship to K.X.Z., and N.I.H grants to B.R.G. (5R01GM067842) and S.L.Z. (5R01DC006485). S.L.Z. is an investigator of Howard Hughes Medical Institute.
VITA

October 4, 1983  Born, Chiba, Japan

2006  B.S., Biophysics and Biochemistry
       University of Tokyo
       Tokyo, Japan

2006-2011  Study Abroad Scholarship
           The Nakajima Foundation
           Tokyo, Japan

2006  Chancellor’s Prize
       University of California, Los Angeles
       Los Angeles, California

2011-2012  Dissertation Year Fellowship
           University of California, Los Angeles
           Los Angeles, California

PUBLICATIONS AND PRESENTATIONS

Probabilistic splicing of Dscam1 establishes identity at the level of single neurons.
Cell in press.

dissection of Dscam alternative splicing dynamics mediated by RNA-binding
proteins. Talk presented at the Cold Spring Harbor Eukaryotic mRNA Processing

Dscam1 splicing in vivo. Poster presented at the Gordon Conference on Neural
Development, Newport, Rhode Island.
Chapter 1

Introduction
The Roles of Cell Recognition Molecules in Wiring of the Nervous System

Neurons that constitute a nervous system must be connected to each other in a precise and specific fashion during the development, in order to construct a functional circuit capable of executing various behaviors that allow the animals to survive. Processing of information relies on the highly ordered architecture of the circuit, and defects in the connections can lead to a broad range of neurological diseases accompanied, in many cases, by behavioral disorders. The assembly of the circuit involves multiple steps, including axon path finding, morphogenesis of dendritic processes, and synaptic partner matching. How these tasks are achieved accurately at the molecular level continues to be a fascinating question in neuroscience.

Half a century ago, Roger Sperry proposed that neuronal processes navigate through the environment using chemical cues to find their targets. Using the optic system of fishes, he analyzed the regeneration patterns of the optic fibers into the tectum after severing the optic trunk. He demonstrated that even when the fibers were scrambled, their regeneration resulted in normal connectivity patterns. He also noted that the route by which these optic fibers reach the terminals was specific. Known as the chemoaffinity hypothesis, he proposed that such precise matching of axons and their targets must be guided by “individual identification
tags, presumably cytochemical in nature, by which [the cells] are distinguished one from another almost, in many regions, to the level of the single neuron” (Sperry, 1963). The hypothesis was met with criticisms, as the prevailing view at the time was that the specificity of the neural patterning originated from orienting effects of mechanical stresses. As Sperry had envisioned, though, studies in the recent decades have established that neurons rely on cell recognition molecules to sense and interact with their environment.

Several cell recognition molecules that guide the assembly of neural circuits have been identified to date. These include members of several different protein families, such as the immunoglobulin superfamily and the cadherins (Dickson, 2002; Tessier-Lavigne and Goodman, 1996). The molecules, many of which are evolutionarily conserved among different species, have been shown to promote adhesion and repulsion of the neuronal processes in both contact-dependent and independent fashion (i.e., the ligands can be presented on the cell surface or secreted). However, the number of genes encoding cell recognition molecules is a fraction of the number of neurons that make up the nervous system in many species and seems much too small to generate enough distinct codes for every neuron. In fact, only about 25,000 genes are encoded in the entire human genome, while some hundred trillion synapses are made by approximately hundred billion neurons (Williams and Herrup, 1988). How, then, could such
complex neuronal connections be specified in a highly ordered fashion? Recent studies have suggested that the circuits may be assembled through the use of a relatively small number of genes in a diverse manner, rather than through a simplistic mechanism where each connection is specified in a matching of “one lock-one key”.

One approach that has been adopted by multiple systems is the use of gradients of cell recognition molecules. Initially proposed by Sperry, this mechanism was first demonstrated in the vertebrate visual system. Here, the graded expression of Eph receptors and their ligands, ephrins, along the nasotemporal (N-T) and dorsoventral (D-V) axes in the retina and along the anteroposterior (A-P) and D-V axes in the tectum, allow the two-dimensional image on the retina to be projected onto higher brain centers with the same spatial topography (Figure 1-1A) (Braisted et al., 1997; Cheng et al., 1995; Drescher et al., 1995; Holash and Pasquale, 1995). This is known as the retinotopic map. Layered on top of these graded expression patterns is the bi-functionality of the ligands, which act both as repellents and attractants in a concentration dependent manner (Hansen et al., 2004; McLaughlin and O’Leary, 2005). This guides the axons to a point on the map where the two opposing forces are balanced. Thus, graded expression of cell recognition molecules allows the formation of topographic maps by
Figure 1-1 I Various Strategies in Neural Circuit Assembly

(A) The retinotopic map in the tectum is generated through a graded expression of Eph-ephrin receptor-ligand pairs. Ephrin-B and its receptor, EphB, form opposing gradients along the dorso-ventral axis in the retina and the tectum, while ephrin-A and EphA form opposing gradients along the perpendicular axis. While ephrins act as attractants in low concentrations, their expression at higher concentrations can lead to repulsion of axons that express only low levels of the Eph receptors. This bifunctionality of the ligands allows the axons to terminate at specific points at which the attractive and repulsive forces are balanced. Thus, the relative coordinates of the neurons in the retina are preserved in the axon terminals in the tectum.

(B) Each olfactory sensory neuron (OSN) expresses a single OR. The levels of ligand-independent G-protein signaling and its downstream cAMP/PKA signaling vary for individual OR, and this is postulated to determine the levels of neuropilin 1 and semaphorin 3A expressed in each neuron. Thus, OSNs expressing the same
organizing a large number of incoming axons along the gradients of just a few molecules.

Guidance of neuronal projections in stepwise fashion is another mechanism by which neuronal connections are specified with a relatively small number of genes. For instance, the axons of olfactory sensory neurons (OSNs) in the vertebrate olfactory system are sorted prior to target innervation through axon-axon interactions mediated by graded expression of Neuropilin 1 and its repulsive ligand, Semaphorin-3A (Imai et al., 2006). This determines their coarse positions along the A-P axis in the target, the olfactory bulb. Once at the target, a different set of classical guidance molecules (e.g., Kirrels) regulate local axon-axon interactions to further sort the neighboring axons into distinct bundles according to the odorant receptor (OR) that each neuron expresses (Figure 1-1B) (Serizawa
et al., 2006). Such stepwise guidance eliminates the necessity to provide a distinct code for every target. This is particularly relevant for the olfactory system, where about a thousand subclasses of OSNs, each specified by the single OR that each neuron expresses, project to distinct target sites in the olfactory bulb. Many systems are now thought to develop through such cooperative actions of cell recognition molecules that provide adhesive and repulsive cues from both long and short ranges (Dickson, 2002; Tessier-Lavigne and Goodman, 1996). Every step effectively reduces the number of potential targets through the use of a relatively small number of cell recognition molecules.

Recently, yet another strategy was uncovered from studies on Dscam1 in Drosophila and the clustered protocadherins in vertebrates. These genes encode large sets of cell recognition molecules with distinct binding affinities from single genes. The discoveries resurrected the intriguing notion that unique identities at the level of single neurons are used for assembling the neural circuit. In fact, the number of cell recognition molecules produced from Dscam1 is so enormous that the gene could potentially endow “individual identification tags” to tens of thousands of neurons. Could these isoforms act as the “chemoaffinity” factors that mediate matching of thousands of synaptic connections?
This dissertation describes our work on \textit{Dscam1}, aimed at understanding how the nervous system utilizes the immense complexity of the gene locus \textit{in vivo}. The results demonstrate that each neuron acquires a unique Dscam1 identity on a probabilistic basis throughout the nervous system. Thus, as Sperry envisioned, neurons indeed acquire “individual identification tags.” Yet, due to the probabilistic nature, these tags are unlikely to contribute to the interactions between different neurons. Rather, together with previous data, our results suggest that they contribute to the assembly of neural circuits through its widespread role in mediating self-recognition of the neurons.

\textit{Drosophila Dscam1} Encodes a Vast Number of Isoforms that Exhibit Isoform-Specific Homophilic Binding

\textit{Dscam1} was isolated in \textit{Drosophila} as an interacting partner of Dock, an adapter protein required for axon guidance (Schmucker et al., 2000). It is one of four \textit{Drosophila} homologs of DSCAM (\textit{Down syndrome cell adhesion molecule}), initially identified on human chromosome 21 as a candidate gene responsible for the neurological phenotypes associated with Down syndrome. Dscam proteins belong to the immunoglobulin (Ig) superfamily, comprising 10 Ig domains (Ig1-Ig10), 6 fibronectin type III repeats, one transmembrane domain, and a cytoplasmic tail. In pancrustaceans, the \textit{Dscam1} locus has expanded
considerably to encode a vast number of splicing isoforms that share the same
domain structure but contain variable Ig domain sequences (Armitage et al.,
2012). This diversification is not observed in the homologs of other groups,
including the vertebrates.

Dscam1 isoforms are generated through mutually exclusive alternative splicing at
multiple exons. In *Drosophila melanogaster*, exons 4, 6, 9, and 17 form clusters
that comprise 12, 48, 33, and 2 alternative variants, respectively (Figure 1-2). The exons correspond to the first half of Ig2, the first half of Ig3, the entire Ig7,
and the transmembrane domain (Schmucker et al., 2000). Splicing at each
cluster is independent of the other, and thus, *Dscam1* has the potential to
generate 19,008 different extracellular domains (i.e. 12 X 48 X 33 = 19,008), tethered to the membrane by one of the two transmembrane domains. In fact,
18,496 of the 19,008 possible extracellular domains have been recently detected
through deep sequencing of mRNA isolated from flies (Sun et al., 2013). The
cytoplasmic domain is also known to generate diversity through skipping at
several exons (S.K.M. and S.L.Z., unpublished results; Yu et al., 2009). The
functional significance of the cytoplasmic domain diversity is not clear.

An extraordinary feature of Dscam1 is the binding specificity of the extracellular
domains. Initially, the specificity was assessed with 11 isoforms, using a series of
Figure 1-2 | Mutually Exclusive Alternative Splicing of Dscam1 Generates a Vast Number of Isoforms That Exhibit Isoform-Specific Homophilic Binding

(A) Dscam1 genomic locus. Exons 4, 6, 9, and 17 form clusters that are alternatively spliced in a mutually exclusive manner. They encode 12, 48, 33, and 2 alternative exons, respectively. Dscam1 proteins comprise 10 immunoglobulin (Ig) domains, 6 fibronectin type III repeats, a single-pass transmembrane domain (TM), and a cytoplasmic region. Exons 4 and 6 correspond to the first half of Ig2 and Ig3, exon 9 corresponds to the entire Ig7, and exon 17 corresponds to the transmembrane domain. As splicing in each cluster is independent of the other, Dscam1 has the potential to generate up to 19,008 isoforms tethered to the membrane by one of the two transmembrane domains.

(B) The variable Ig domains determine the binding specificity of the isoforms. Biochemistry and X-ray crystallography has shown that the isoforms exhibit isoform-specific homophilic binding and that each variable Ig domain interacts with the same domain on the opposing molecule in a modular fashion. Thus, all three variable Ig domains must match to achieve binding (top panel). A difference in one Ig domain results in very little to no binding (bottom panel).
in vitro bead aggregation assays and biochemistry (Wojtowicz et al., 2004). The experiments showed that all the isoforms tested engage in isoform-specific homophilic binding. In contrast, little to no heterophilic binding was observed, even for a pair of closely related isoforms that differ by just seven to twelve amino acids in one of the variable Ig domains. Based on these findings, it was proposed that each variable Ig domain binds to the identical variable Ig domain in an opposing molecule in a modular fashion, and that isoform-specific homophilic binding was a general property of Dscam1 isoforms.

To test the binding specificity in a more extensive manner, Wojtowicz et al. developed a high-throughput ELISA based binding assay (Wojtowicz et al., 2007). As the previous data suggested that the binding was modular (and later confirmed in the crystal structure (Sawaya et al., 2008)), it was reasoned that homophilic binding of the Dscam1 isoforms can be tested by assessing the binding of each of Dscam1’s 93 different variable Ig domains (i.e., 12 Ig2 variants, 48 Ig3 variants, and 33 Ig7 variants) to an identical variant in an opposing molecule. The results argued that more than 18,000 isoforms (>95% of all isoforms) exhibit isoform-specific homophilic binding, with little to no heterophilic affinity. From analytical centrifugation performed on a few isoforms, the $K_d$ values of monomeric binding were later determined to be between 1 and 14 $\mu$M, while the binding between two isoforms that differ in only a single Ig2 domain were
below the detection limit (i.e., >500 µM) (Zipursky and Grueber, 2013; Wu et al., 2012).

**Multiple Dscam1 Isoforms Are Expressed in a Single neuron**

Early attempts to analyze the splicing patterns of the Dscam1 extracellular domain in neurons were performed by using customized oligonucleotide microarrays on neuronal tissues and on populations or single cells isolated by fluorescence-activated cell sorting (FACS) (Neves et al., 2004; Zhan et al., 2004). Prior to the work by Chess and his colleagues, it was not known whether individual cells expressed one or multiple isoforms, and whether there was any cell-type specificity to the splicing pattern. To assess whether the splicing pattern was differentially regulated across different neuronal populations, the entire population of photoreceptor neurons, a couple of distinct classes of photoreceptor neurons, and the eye-antennal disk isolated from third instar larvae were analyzed. These different populations all expressed many different variants from all three clusters of alternative exons (i.e., exons 4, 6, and 9), arguing that cells of a specific cell type collectively express multiple isoforms. A comparison between R3/R4 and R7 photoreceptor populations also revealed that although the spectrum of spliced variants was broad in each population, it was differentially biased with regards to exon 9 (Neves et al., 2004). No significant difference was
observed for exons 4 and 6. Similar microarray analyses performed by Zhan et al. on mushroom body (MB) neurons of the central brain also showed a bias in splicing of exon 9, which was also different from that of the photoreceptors (Zhan et al., 2004). Thus, these results showed that the spectrum of Dscam1 isoforms expressed in each class of neurons was distinct, though with some overlaps.

Microarray analyses following single-cell RT-PCR on a limited number of isolated photoreceptor neurons (i.e., seven R3/R4 neurons and seven R7 neurons) revealed that each cell expressed multiple alternative variants of exon 9. Importantly, individual neurons were found to express different combinations of the alternative variants. Through computer-based simulations, it was estimated that each neuron expressed 14-50 isoforms. Based on these results, Chess and colleagues proposed that each cell expressed a set of Dscam1 isoforms in a stochastic fashion. Although it is not clear whether the photoreceptors require Dscam1 for their proper development, Zhan et al. have also made a similar observation in isolated MB neurons that require Dscam1 (Zhan et al., 2004). These studies raised an intriguing possibility that each neuron acquired a unique cell surface identity through a combinatorial expression of Dscam1 isoforms. Whether this was the case throughout the nervous system, and if so, how this could be achieved was not addressed.
Dscam1 Promotes Self-Avoidance in Neurons

How do Dscam1 isoforms contribute to the development of neural circuits? The first indication of their function came from genetic studies performed on the intrinsic neurons of the MB, the Kenyon cells (Wang et al., 2002; Zhan et al., 2004). The MB is a bi-lobed structure in the central brain and comprises ~2,500 intrinsic neurons, called Kenyon cells. The axon of each Kenyon cell bifurcates at a specific branching point, each giving rise to two sister branches. In wildtype animals, the two branches extend into separate MB lobes. In contrast, when single Dscam1 mutant cells were generated in an otherwise wildtype background, the sister branches often failed to segregate into their respective lobes and instead extended into the same lobe (Figure 1-3A). Importantly, the null phenotype in single cells could be rescued with overexpression of a single arbitrarily chosen isoform, suggesting the lack of isoform-specific function. Together with the biochemical evidence that the isoforms exhibit isoform-specific homophilic binding and that each neuron expresses multiple isoforms, the data led to the idea that each neuron acquires a unique combination of Dscam1 isoforms and that the homophilic binding of Dscam1 isoforms triggers a repulsive interaction specifically between the sister branches, a phenomenon known as self-avoidance.
Figure 1-3 | Dscam1 Mediates Self-Avoidance

(A) Dscam1 mediates repulsion of the sister branches of Kenyon cells, intrinsic neurons of the MB. Each Kenyon cell sends an axon through the peduncle (P), after which it bifurcates into two sister branches. In wildtype cells, the sister branches segregate and grow in different directions, giving rise to the bi-lobed structure of the MB. In a Dscam1 mutant cell in an otherwise wildtype background, the sister branches fail to segregate and enter the same lobe (red neuron, right panel).

(B) In the sensory neurons of Drosophila larvae, Dscam1 promotes uniform coverage of their receptive fields through self-avoidance. Here, class III da neuron (green) and class I da neuron (black) with overlapping dendritic fields are shown. Both neurons exhibit self-avoidance under wildtype conditions (left panel). In a Dscam1 mutant neuron in an otherwise wildtype background, its sister branches fail to segregate and cross each other extensively (middle panel). Overexpression of the same single Dscam1 isoform in two neighboring neurons cause their dendritic branches to repel each other, resulting in mutually exclusive receptive fields (right panel). Thus, this indicates that neighboring neurons must express different sets of Dscam1 isoforms.
The notion that Dscam1 mediates self-avoidance was demonstrated conclusively in the sensory neurons of the peripheral nervous system (PNS) in the *Drosophila* larvae (Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007). These sensory neurons are called the dendritic arborization (da) neurons, and they are classified into four classes (class I-IV) that exhibit varying levels of dendritic complexity. Their dendritic receptive fields innervate the body wall in a planar fashion, making them an ideal system to study interactions between processes of both the same and neighboring neurons. Each segment contains a stereotypic set of the da neurons. The dendritic fields of da neurons of the same class typically do not overlap, a phenomenon referred to as tiling. In contrast, different classes of da neurons can overlap, such that the dendritic branches from one neuron grow across those of the other neurons. Importantly, all four classes of neurons exhibit self-avoidance, and the sister branches rarely come into contact with each other. These features allow each sensory neuron to occupy its dendritic field independent of the other types of sensory neurons, while avoiding redundant coverage of the same area by two neurons of the same type.

In single neurons mutant for *Dscam1*, sister branches failed to segregate and often crossed each other and fasciculated (Figure 1-3B), consistent with the role of Dscam1 in contact dependent repulsion (Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007). This resulted in a highly non-uniform receptive field.
Importantly, the lengths and the number of branches remained unaffected in *Dscam1* null. Similar phenotypes were observed in all four classes of da neurons, indicating that self-avoidance through Dscam1 is a common mechanism in different cell types that share receptive fields. The branch segregation phenotype could be rescued by overexpression of a single arbitrarily chosen wildtype isoform in the mutant neuron, consistent with the experiments in the MB. Interestingly, deletion of the Dscam1 cytoplasmic domain from the rescue construct resulted in adhesion of the dendrites rather than repulsion. Thus, these observations support a model where homophilic binding of Dscam1 isoforms on sister branches trigger downstream signaling that causes them to repel each other.

Additional support for the self-avoidance model came from gain of function studies. Ectopic expression of the same Dscam1 isoform in two neighboring neurons that normally share their receptive fields resulted in the two neurons avoiding each other, forming mutually exclusive dendritic fields (Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007). Together with the observation that multiple Dscam1 isoforms are expressed in each neuron, these data demonstrate that each neuron must acquire a unique Dscam1 identity to distinguish between self and non-self. Moreover, no particular isoform is required
for self-avoidance, and it is only important that the neighboring neurons express different isoforms.

A series of genetic experiments were conducted by Wu et al. to critically address whether the specific recognition between Dscam1 isoforms on sister branches is required for self-avoidance in vivo (Wu et al., 2012). Based on the previous structural analyses and through comparisons of amino acid sequences from Dscam1 homologs in 39 species, pairs of chimeric isoforms with newly engineered Ig2 domains were generated that had lost homophilic binding but had simultaneously gained heterophilic binding specificity to each other. These chimeric isoforms were knocked into the endogenous locus, and their ability to promote self-avoidance was tested in the MB and the da neurons. While single wildtype isoforms expressed in a neuron were capable of supporting self-avoidance, the expression of single chimeric isoforms showed reduced ability to support self-avoidance. By contrast, co-expression of two complementary chimeric isoforms in a single MB neuron restored self-avoidance. Further, co-expression of these isoforms in neighboring da neurons with overlapping receptive fields caused ectopic repulsion between their dendrites, resulting in mutually exclusive receptive fields. These results demonstrate that homophilic binding of Dscam1 isoforms on opposing surfaces of sister neurites is required for self-avoidance.
Thousands of Dscam1 Isoforms Are Required for Proper Development

Although the experiments described so far suggest that a single arbitrarily chosen Dscam1 isoform is sufficient to promote self-avoidance in a single neuron, diversity of the isoforms is essential to support self-avoidance in many neurons while simultaneously allowing their co-existence. Indeed, as previously noted, genetic evidence suggests that each neuron must express a unique combination of Dscam1 isoforms to avoid ectopic repulsion (Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007). Thus, the diversity of Dscam1 isoforms must be large enough to accommodate a vast number of neurons that interact during development.

How large does the Dscam1 isoform diversity need to be, in order to support self-avoidance in multiple neurons in vivo? The number of isoforms required for the robust discrimination between self and non-self was addressed genetically by reducing the number of isoforms (Hattori et al., 2009). By homologous recombination, either both exon clusters 6 and 9 or exon cluster 9 alone were fixed with single variants through the replacement of exons 5-11 or 7-11 with cDNAs, respectively. These alleles limited the number of potential isoforms to 12 or 576. By combining two alleles with different fixed variants, animals that encode
24 or 1,152 isoforms were also generated. In addition, deletion alleles were analyzed in which either 3 or 9 alternative variants were removed from the exon 4 cluster, limiting the number of potential isoforms to 14,256 and 4,752, respectively. The alleles with reduced number of isoforms were then tested for their ability to support discrimination between self and non-self in the da neurons and the MB.

Self-avoidance in single neurons was supported with an animal encoding just 12 isoforms, consistent with previous reports that a single isoform is sufficient for the process. In contrast, animals with 1,152 isoforms exhibited defects in discrimination between self and non-self in the da neurons, resulting in ectopic repulsion between two classes of da neurons that normally overlap. They also exhibited defects in development of the bi-lobed MB structure. In contrast, both the da neurons and the MB were indistinguishable from the wildtype in the deletion mutant encoding 4,752 isoforms. Thus, thousands of isoforms were required to ensure unique Dscam1 identity in each neuron.

**Analysis of Dscam1 Splicing: Deterministic or Probabilistic?**

Genetic and biochemical evidence establish a strong case for Dscam1 in mediating self-avoidance, by endowing each neuron with a unique cell surface
identity that distinguishes it from the other neurons. On the other hand, expression studies had been mostly limited to microarrays and other *in vitro* methods performed on populations of neurons or a handful of single isolated neurons from just a few cell types. Thus, although unique Dscam1 expression in each neuron is a critical feature during development, how this is achieved had not been addressed. Moreover, an important question still remained as to whether Dscam1 could mediate interactions between different neurons, for example to match synaptic connections. Chapter 2 describes our work that was aimed to answer these questions and comprises a version of a manuscript by Miura et al., currently in press in *Cell*. Here, we took a genetic approach to visualize the splicing pattern of *Dscam1* throughout the animal. This enabled analyses of the splicing within an animal and among multiple animals in both fixed tissues and live preparations. The results led us to conclude that the splicing of *Dscam1* is probabilistic. No deterministic splicing patterns, such as cell-type or positional specific patterns, were observed throughout the nervous system. In the final chapter, I discuss the broader implications of our results and some unresolved issues relating to the probabilistic splicing, as well as the evolution of genes that mediate self-avoidance in different animals.
References


Chapter 2

Probabilistic Splicing of *Dscam1* Establishes

Identity at the Level of Single Neurons
Abstract

The *Drosophila Dscam1* gene encodes a vast number of cell recognition molecules through alternative splicing. These exhibit isoform-specific homophilic binding and regulate self-avoidance, the tendency of neurites from the same cell to repel one another. Genetic experiments indicate that different cells must express different isoforms. How this is achieved is not known, as the expression of alternative exons *in vivo* has not been shown. Here, we modified the endogenous *Dscam1* locus to generate splicing reporters for all variants of exon 4. We demonstrate that splicing does not occur in a cell-type specific fashion, that cells sharing the same anatomical location in different individuals express different exon 4 variants, and that the splicing pattern in a given neuron can change over time. We conclude that splicing is probabilistic. This is compatible with a widespread role in neural circuit assembly through self-avoidance and is incompatible with models in which specific isoforms of Dscam1 mediate homophilic recognition between processes of different cells.
Introduction

Neural circuits are assembled through interactions between neurites, both axons and dendrites, of vast numbers of neurons. This assembly relies upon many different receptors and ligands mediating repulsive and adhesive interactions between neurites. Recent studies have highlighted the importance of repulsive interactions between neurites of the same cell for patternning neural circuits. This process, first described in the leech and termed self-avoidance (Kramer and Stent, 1985), contributes to circuit assembly in both vertebrates and invertebrates (Hattori et al., 2007; Hughes et al., 2007; Lefebvre et al., 2012; Matthews et al., 2007; Millard et al., 2010; Soba et al., 2007; Wang et al., 2002a). Self-avoidance relies on neurites acquiring a cell surface identity specific to each neuron, different from other neurons they encounter during development (Schmucker and Chen, 2009; Zipursky and Grueber, 2013; Zipursky and Sanes, 2010).

Self-avoidance is understood best in Drosophila where a large family of immunoglobulin (Ig) containing proteins encoded by the Dscam1 locus mediates this process (Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007; Wang et al., 2002a; Zhan et al., 2004). Dscam1 proteins exhibit isoform-specific homophilic binding, both in vitro and in vivo (Wojtowicz et al., 2004; Wojtowicz et al., 2007; Wu et al., 2012). Upon contact between neurites of the same cell,
homophilic binding of Dscam1 triggers repulsion (Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007; Wu et al., 2012). The importance of Dscam1 for self-avoidance has been demonstrated in axons, dendrites, and postsynaptic elements at multiple contact synapses (Hattori et al., 2007; Hughes et al., 2007; Matthews et al., 2007; Millard et al., 2010; Soba et al., 2007; Wang et al., 2002a; Zhan et al., 2004). Genetic studies indicate that thousands of isoforms are necessary for self-avoidance and neurons must express different Dscam1 isoforms from their neighbors for normal patterning (Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007). As neurons often encounter the neurites of many different cells, particularly in dense neuropil in the developing central nervous system (CNS), robust mechanisms must exist to ensure that neurons that encounter each other during development express different isoforms.

Dscam1 isoforms are generated through alternative splicing (Schmucker et al., 2000). Each extracellular domain shares the same overall domain structure, but differs in amino acid sequence in one or more of three variable Ig domains, designated Ig2, Ig3, and Ig7. These are encoded by clusters of alternative exons with 12 variants of exon 4 (encoding half of Ig2), 48 variants of exon 6 (encoding half of Ig3), and 33 variants of exon 9 (encoding Ig7 in its entirety) (Figure 2-1A). Each combination of three variable domains determines the unique binding specificity of each isoform. Indeed, some 18,000 of the 19,008 potential
extracellular domains exhibit strong isoform-specific homophilic binding (Wojtowicz et al., 2004; Wojtowicz et al., 2007). The ability of neurites to discriminate between self and non-self depends critically on the pattern of alternative splicing of Dscam1 in each neuron.

Although genetic studies indicate that neurons with overlapping dendrites and axons must express different isoforms, the expression of Dscam1 isoforms in vivo has not been described. Microarray studies on cDNAs prepared by RT-PCR from a small number of single R7 and R3/R4 photoreceptor neurons, isolated by FACS, indicated that a single neuron expresses multiple variants of exon 9 (Neves et al., 2004). Although it remains unclear whether photoreceptor neurons require Dscam1 for circuit assembly, using the same molecular approach we later reported a similar mode of expression in a small number of mushroom body (MB) neurons, which require Dscam1 for self-avoidance (Zhan et al., 2004). These findings indicate that photoreceptor and MB neurons express different combinations of Dscam1 isoforms conferring unique cell identities to them.

It has not been demonstrated how the expression of a unique combination of Dscam1 isoforms in each neuron is determined. Neurons may express Dscam1 isoforms in a highly regulated or deterministic fashion reflecting their cell type, their unique location (e.g. their dorsoventral or anteroposterior location),
developmental history (e.g. their birthdate or lineage) or some combination of such developmental determinants. Alternatively, the pattern of isoform expression may not be regulated in a deterministic fashion, but rather may result from lack of regulation. A probabilistic choice of isoforms, in theory, provides a robust and efficient means by which neurons acquire unique identities (Forbes et al., 2011; Hattori et al., 2009). Furthermore, different strategies may be employed in different systems. Distinguishing between these possibilities has been hampered by the difficulties in assessing the expression of alternative isoforms at the level of identified neurons within the developing nervous system.

Here we report that Dscam1 splicing is probabilistic through the analysis of alternative exon 4 expression using splicing reporters. Live imaging revealed that the alternative variants spliced in a given cell also change over time. A similar mode of splicing was observed in cells requiring Dscam1 for self-avoidance in axons, dendrites, and postsynaptic elements, and more generally throughout the nervous system. These findings support the notion that the ability of neurites to discriminate between self and non-self relies on a probabilistic mechanism to endow developing neurons with unique cell surface identities. As specific cell types were not found to reproducibly splice the same alternative exon 4, and thus they do not express isoforms with the same binding specificity, it is unlikely that neurons use Dscam1 to specify interactions between different neuronal cell types.
Conversely, widespread probabilistic splicing supports a role for Dscam1 diversity in self-recognition throughout the developing nervous system.

Results

Modification of the Endogenous Locus to Detect Expression of Each Alternative Variant of Exon 4

We devised a splicing reporter system to visualize the expression of single alternative variants of exon 4 from the endogenous Dscam1 locus in individual cells in vivo. Reporter constructs were knocked into the endogenous locus by homologous recombination of exons 3-5 with a replaceable cassette, followed by phiC31 recombination mediated cassette exchange (RMCE; see Figure 2-S.1 and Extended Experimental Procedures). The detection system was designed to monitor expression in fixed preparations and in live animals at different time points, and to identify cells expressing different alternative exons by their morphology, co-expression of specific markers, or both.

We generated reporters for each of the 12 alternative variants of exon 4. Each variant is referred to as exon 4.X, where X indicates the position of the alternative variant within the cluster from 5’ to 3’ (i.e. exon 4.5 is the fifth alternative exon
Figure 2-1 | The Design of Reporters for Splicing of Alternative Exon 4 Variants

(A) Schematic representation of the Dscam1 genomic locus. Color-coded exons are alternatively spliced in a mutually exclusive manner, such that one variant each from exon 4, exon 6, exon 9, and exon 17 clusters are included in the mature mRNA. Exons 4, 6, and 9 correspond to three variable Ig domains in the extracellular domain that determine the binding specificity of the isoform and encode 12, 48, and 33 variants, respectively. Thus, this gene has the potential to generate up to 19,008 distinct extracellular domains.

(B) Splicing reporter design. The exon 4.5 reporter is shown as an example. All variants of exon 4, except exon 4.5, were mutated by a single base pair insertion. A transmembrane domain (TM), "self-cleaving" 2A peptide, and Gal4 followed by a stop codon and polyA site were fused in frame to exon 5. Splicing of exon 4.5 results in the translation of Gal4, which drives the expression of GFP markers under the control of UAS elements. Splicing of any other alternative exon 4 results in a frame shift, generating a stop codon in exon 5. Reporters were generated for all 12 alternative variants. A positive control with all wild-type exon 4 variants and a
negative control with all exon 4 variants mutated were also generated. For detailed experimental strategy of knock-in generation, see Experimental Procedures and Figure 2-S.1.

downstream from exon 3). In each reporter, only a single alternative variant was in-frame with a downstream indicator, the transcriptional activator GAL4. The remaining 11 alternative variants all included a frame-shift mutation that rendered the indicator out-of-frame (Figure 2-1B). Thus, GAL4 was expressed only in neurons that spliced the single remaining wild type alternative variant. These neurons were visualized by expression of either membrane-bound or nuclear GFP under the control of tandemly arranged UAS elements. Positive and negative control alleles were also generated, in which all variants were in frame ($Dscam1^{positive-Gal4}$) or out of frame ($Dscam1^{negative-Gal4}$). As the reporter knock-in mutations inactivate $Dscam1$, all experiments were done in a heterozygous background. No morphological or developmental defects were observed in this background in any of the systems we analyzed. As expected, positive controls were expressed widely, and only very few weakly stained scattered cells were observed in the negative control. By contrast, each of the 12 single exon reporters was expressed in a salt-and-pepper pattern within the developing nervous system. Here, we describe the expression patterns of all 12 variants of
alternative exon 4 in developing MB neurons in the central brain, dendritic arborization (da) neurons in the periphery, and visual system neurons.

**MB Neurons Express Each Variant of Alternative Exon 4 at a Characteristic Level and Frequency**

A role for Dscam1 in self-avoidance was first described in the MB (Zhan et al., 2004). MB is a bi-lobed structure in the central brain, with each lobe containing one of two sister branches of intrinsic MB neurons (i.e. Kenyon cells) (Figure 2-2A). Each MB contains about 2,500 Kenyon cells. These cells send axons within a common nerve called the peduncle. Axons bifurcate at the base of the peduncle, and the branches segregate and extend into separate lobes with high fidelity. In the absence of Dscam1, as assessed in single mutant cells, sister branches often fail to segregate, and instead, the two branches extend within the same lobe. Here, we argued that axons from different neurons express different Dscam1 isoforms, such that at the common branch point branches discriminate between self and non-self. This “self-recognition” would lead to homophilic repulsion and branch segregation (Hattori et al., 2007; Wang et al., 2002a; Zhan et al., 2004).
Figure 2-2 I Alternative Variants of Exon 4 Are Expressed at Different Frequencies in the MB
Membrane-bound GFP tagged with a V5 epitope (white) was used as the readout of the splicing reporters in the MB lobes (blue) and their intrinsic neurons, Kenyon cells (blue), of mid-pupal brains (65 hr after puparium formation). MB lobes and Kenyon cells were visualized by anti-Fasciclin II and anti-Dachshund, respectively.

(A) Schematic of the MB. Kenyon cell bodies form a cluster in the posterior part of the brain. Each Kenyon cell sends an axon through the peduncle (P). Each axon bifurcates and the branches extend into two different lobes. This segregation of the sister branches requires repulsion induced by homophilic binding of the Dscam1 isoforms. Before entering the peduncle, Kenyon cells also form a dendritic field in a structure called the calyx (C).
We assessed expression of all 12 variants of exon 4 during MB development in pupae. \textit{Dscam}^{1\text{positive-Gal4}} was expressed strongly in both lobes of the MB (Figure 2-2C1). This resulted from the reporter expression in the Kenyon cells and possibly other neurons sending process into the lobes (i.e. MB extrinsic neurons). Most, if not all, Kenyon cells expressed \textit{Dscam}^{1\text{positive-Gal4}} (Figure 2-2C2), consistent with previous data indicating Dscam1 expression in these cells (Zhan et al., 2004). By contrast, only a couple of weakly stained cells were observed in the negative control (i.e., \textit{Dscam}^{1\text{negative-Gal4}}) (Figures 2-2B1 and 2-2B2). The source of this background expression is not known.

Each of the 12 exon 4 variants was expressed in a subset of cells (Figures 2-2D - 2-2G and Figures 2-S.2A - 2-S.2J). For a given alternative variant, the expression level and number of cells expressing the variant were qualitatively similar between animals. By contrast, the expression frequency for different alternative
variants was different (Figures 2-2D - 2-2G and Figures 2-S.2A – 2-S.2J). For example, the frequency of exon 4.2 splicing was consistently lower than any other alternative variant we analyzed (Figures 2-2E₁ and 2-2E₂). In contrast, exon 4.12 was expressed in many more cells at much higher levels (Figures 2-2G₁ and 2-2G₂). Furthermore, although the frequency of expression of a given variant was similar between animals, there was no obvious similarity in the distribution of cells in different animals. The overall bias in splicing of the alternative variants is consistent with our previous expression data obtained on populations of Kenyon cells isolated by FACS (Zhan et al., 2004). For example, microarray and cDNA sequencing results both indicated that exons 4.2 and 4.12 represent the two extremes of the frequency spectrum with exon 4.2 expressing the least and exon 4.12 the most. The convergence of these data argues that the splicing trap method accurately reflects splicing in vivo.

**Splicing in Single Class IV da Neurons Is Probabilistic**

To assess rigorously whether *Dscam1* alternative splicing is probabilistic, we analyzed expression in class IV da neurons. The dendrites of these neurons require Dscam1 for self-avoidance. In wild type, each class IV da neuron forms highly branched dendrites that do not overlap (Grueber et al., 2002). By contrast, in the absence of Dscam1, these dendrites frequently cross one another, form
clumps, and as a consequence disrupt coverage of the receptive field (Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007).

Each class IV da neuron can be identified reliably from animal to animal based on its unique position, morphology, and marker expression. There is a dorsal (ddaC), lateral (v’ada), and ventral (vdaB) class IV neuron in each abdominal hemisegment A1 through A7 (Grueber et al., 2002; Han et al., 2012) (Figure 2-3A). Thus, this system allows us to score the alternative splicing of exon 4 in 42 different uniquely identified neurons in each animal, and to compare expression between these cells and between the same cell in different animals.

We first set out to address whether the choice of alternative exon 4 variant is always the same for a class IV neuron in the same location in different animals. GFP-tagged histone H2A was used to detect the expression of each splicing reporter, and class IV da neurons were identified by the expression of tandem dimer Tomato under the control of a genomic element from the pickpocket locus (Han et al., 2011) (Figures 2-3B and 2-3C). Each neuron in abdominal segments A1 through A7 was identified, and its nuclear GFP intensity was quantified to determine the expression (i.e., simply ON or OFF). GFP intensity in Dscam$^\text{negative-Gal4}$ was used to determine the threshold between ON and OFF. Briefly, the histogram was fit to a Gamma probability distribution function to
determine a threshold value; 312/314 of the control cells were below this value (for details, see Extended Experimental Procedures). With this threshold, the positive control was expressed in all neurons scored except one (i.e., 1/299), confirming Dscam1 expression at this stage (Figure 2-3E). By contrast, in every animal, only a subset of the 42 class IV da neurons expressed each alternative
exon 4 variant (Figures 2-3F - 2-3I and Figures 2-S.3A - 2-S.3G). Moreover, the set of neurons that spliced a given variant was different between individual animals, and no cell expressed the same variant across all animals, with the exception of 3 cells that expressed exon 4.2 in all 8 animals analyzed (Figures 2-3F - 2-3J and Figures 2-S.3A - 2-S.3G). Given that exon 4.2 is expressed in ~80% of class IV neurons, we would expect on average 7 cells to express this alternative exon in all animals by chance. Thus, the splicing choice is not deterministic.

To determine whether the probability of splicing a specific variant is the same for each class IV da neuron (i.e., whether there is a splicing bias for cells in certain locations), we performed statistical simulations. For a given alternative exon 4 variant, the spatial splicing patterns were shuffled to generate a set of randomized patterns. The pairwise Pearson correlation coefficient (PCC) was then calculated for every pair within the set, and the average score was determined. This process was repeated a thousand times to generate a distribution of the average PCC of the randomized sets and fit to a normal distribution (Figure 2-S.4A). The differences between the distribution of average PCCs from the simulations and the experimental data for each exon 4 variant were not statistically significant (p>0.025; two-tailed test) (Figures 2-4 and 2-S.4B).
These data are consistent with the notion that the splicing probability of each exon 4 variant is equivalent among class IV da neurons and that the spatial pattern of splicing within the class of neurons is the result of a random process. Thus, the spatial location of a neuron does not contribute to the splicing choice.

In a separate series of experiments, we assessed the splicing of alternative variants of exon 4 in ventral class III (vdaD) and class IV neurons, which share overlapping receptive fields. There was no correlation in splicing with each cell expressing alternative variants independent of the other cell (Figure 2-S.3I). Thus,

Figure 2-4 | Expression of Exon 4 Variants Is Randomly Distributed Among the 42 Class IV da Neurons

Statistical test results for exons 4.1, 4.2, 4.9, and 4.12, indicating no significant difference between the average Pearson correlation coefficients of experimental data (red line) and one thousand trials of randomized patterns (shown as histograms in blue) (p>0.025; two-tailed test). This result was similar for each of the remaining eight variants (see Figure 2-S.4). Thus, although the frequency of expression of different exons varies, the spatial expression pattern is not biased towards any cell and is random. For methods, see text and Figure 2-S.4A.
we conclude that the selection of the specific exon 4 variant to include within a transcript is made probabilistically on a cell-by-cell basis, such that each alternative variant displays varying frequencies of expression (i.e. between ~6-80% for class IV da neurons) (Figure 2-3K).

The Frequency of Inclusion of Exon 4 Variants Is Different Between Cell Types

Cell-type specific differences were observed in the frequency of exon 4 variant splicing in class IV da neurons and in Kenyon cells. While the likelihood that a given alternative variant was expressed in class IV da neurons was consistent from animal to animal, there were significant differences in the percentage of cells expressing different alternatives (Figure 2-3K and Figure 2-S.3H). This distribution was strikingly different from those observed in the Kenyon cells. For instance, exon 4.2 was expressed in ~78% of the class IV da neurons, yet it was only rarely expressed in Kenyon cells (Figures 2-2E, 2-3G, and 2-3K). By contrast, exon 4.1 was expressed in only ~6% of the class IV da neurons, while it was expressed in many more Kenyon cells than exon 4.2 (Figures 2-2D, 2-3F, and 2-3K). These data demonstrate that the probability of splicing a given alternative exon 4 varies between cell types that require Dscam1 for self-avoidance. This may reflect differences in splicing factors expressed in cells or
the level of their expression, as suggested previously by Chess and colleagues (Neves et al., 2004).

**The Splicing Choice of Alternative Exon 4 Variant Changes Over Time**

We sought to address whether a set of alternatively spliced transcripts is generated early in da neuron development and maintained, or whether different transcripts are expressed at different times. To this end, we analyzed splicing of exon 4 variants in the same set of dorsal class IV da neurons ~48 hours apart, imaged in late second and wandering third instar larval stages. Splicing was visualized in reporters for exons 4.6 and 4.10, using GFP-tagged histone-H2A. These were chosen arbitrarily among the variants that exhibited intermediate levels of expression. We assessed whether any neuron switched exon 4.6 or 4.10 expression from OFF to ON or ON to OFF during this time window. *Dscam1*\(^{negative-Gal4}\) animals were used to determine the threshold between ON and OFF (Figure 2-5A). With the positive control, we confirmed that Dscam1 was expressed in all class IV neurons at these developmental stages (Figure 2-5B). This is consistent with a previous study indicating that these neurons continue to arborize during this time period (Grueber et al., 2003).
Figure 2-5 | Splicing of Exon 4 in Class IV da Neurons is Dynamic

(A-D) Expression of exon 4.6 and 4.10 (and controls) was assessed in the dorsal 14 class IV da neurons of second instar larvae, and the expression in these same cells was determined 48 hours later. Thresholds for expression were determined by fitting a Gamma probability density function to the values from negative control animals (for details, see Extended Experimental Procedures). Neurons whose expression changed between the two time points are indicated with asterisks. Color codes are as in Figure 2-3. If the expression in neurons could not be quantified at either of the two stages, they were not included in the analysis and are indicated as grey boxes. Cells are indicated by their positions in each row, and each column represents the cells scored in each animal.

(E) Example of a neuron switching exon 4.6 expression from OFF to ON. White, GFP-tagged H2A driven by exon 4.6 reporter; red dashed circles, locations of nuclei; and blue, class IV specific marker. Signals were saturated in this panel post-acquisition for viewing purposes. Scale bar, 2μm.
The pattern of exon 4.6 and 4.10 changed over time. Seventy-five neurons were identified that initially lacked exon 4.6 expression, and of these, 11 neurons switched to ON (Figures 2-5C, 2-5E, and 2-5F). Conversely, seventeen neurons were identified that initially expressed exon 4.6, and of these, 3 neurons switched to OFF (Figures 2-5C and 2-5F). Thus, one in every 6 to 7 neurons (~15%) changed splicing of exon 4.6. OFF to ON changes were also observed with exon 4.10, although less frequently; ON to OFF changes were not statistically significant (Figures 2-5D and 2-5F). For both exons 4.6 and 4.10, the number of ON to OFF cells may be an underestimate, as GFP tagged H2A and Gal4 proteins could persist well after the reporter mRNA. Based on the analysis of these two variants, we estimate that the splicing of each variant changes in ~10% of the cells (i.e. from ON to OFF and from OFF to ON). By extrapolating these data to all 12 alternative variants of exon 4, many class IV da neurons are likely to change the exon 4 variants spliced during this 48h period. As the positive control indicates that some variant of exon 4 is expressed in all neurons.
throughout the imaging period, our results demonstrate that the choice of alternative exon, and thus the isoforms expressed in a cell, can change over time.

The Pattern of *Dscam1* Splicing in the Visual System Is Consistent with a Broad Role in Self-recognition in the CNS

In the previous sections, we demonstrated that *Dscam1* is spliced in a probabilistic fashion in neurons utilizing Dscam1 for axon and dendritic self-avoidance. In addition to these functions, Dscam1 regulates synaptic organization at tetrad synapses in the visual system through self-avoidance (Millard et al., 2010). Tetrad synapses are multiple contact synapses between photoreceptor axons and the dendrites of target cells called lamina neurons. Each tetrad contains a single presynaptic release site and an invariant pair of postsynaptic elements, one from an L1 lamina neuron and the other from L2 (Figures 2-6A - 2-6A’). Each L1 and L2 neuron contributes postsynaptic elements to multiple tetrads along the same photoreceptor cell axon, and each photoreceptor neuron is presynaptic to only one L1 and to only one L2 neuron. Two additional dendrites from other cells complete the tetrad. While L1/L1 and L2/L2 pairs are not seen at wild type tetrads, in the absence of both Dscam1 and Dscam2 (a paralog expressing two isoforms), many L1/L1 and L2/L2 pairs were observed. We argued that L1 and L2 express different isoforms of Dscam1 and
Figure 2-6 | Scattered Neurons of Specific Cell Types Express Alternative Variants of Exon 4 in the Visual System

(A-A'') Schematic representation of the Drosophila visual system indicating the relationship between the retina (Re), lamina (La), and medulla (Me) (A). The photoreceptors (R), a few classes of lamina monopolar neurons including the L1 and L2 neurons, and two medulla neurons (medulla intrinsic neuron (Mi) and transmedullary neuron (Tm)) are shown. This is only a small subset of the greater than 60 cell types innervating these structures. All neuronal subtypes shown are repeated in the medulla. As such, if specific alternative versions of exon 4 were expressed in each cell of any of these cell types, or others of similar periodicity, highly regular columnar structures and uniform layers would be seen (see Figure 2-S.5A). Such patterns were not observed. Importantly, L1 and L2 require Dscam1 for normal patterning of tetrad synapses via self-avoidance (see text). Each tetrad
Dscam2. This would prevent pairing of postsynaptic elements from the same cell through homophilic repulsion, and, in this way, this mechanism contributes to the appropriate pairing of L1 and L2. We sought to determine whether L1 and L2 express different Dscam1 isoforms, and, if they do, how this is regulated.

Most, if not all, lamina neurons express $Dscam^{positive-Gal4}$ as tetrad synapses are forming within the developing lamina during mid-pupal development, and none
expressed $Dscam^{\text{negative-Gal4}}$ (Figures 2-6B and 2-6C). These findings are consistent with protein expression studies indicating that Dscam1 proteins are expressed on the dendrites of L1 and L2 neurons during this developmental time period (Millard et al., 2010). Although the morphologies of L1 and L2 neurons are very similar, these neurons can be distinguished from one another by the selective expression of the transcription factor Seven-up in L1 (Figure 2-6A) (Claude Desplan, personal communication). All the individual splicing reporters were expressed in subsets of L1 and L2 neurons scattered throughout the lamina with no apparent pattern (Figures 2-6D$_1$ - 2-6G$_1$ and Figures 2-S.5B$_1$ - 2-S.5I$_1$; yellow arrowheads and arrows). Occasionally, these neurons were in the same cartridge, as one would expect from a probabilistic splicing. Presumably, these differ from one another through the expression of additional alternative versions of exon 4 (as in class IV da neurons (see Discussion)), through differential expression of alternative versions from the two other clusters of alternative exons encoding variable recognition domains, or through both mechanisms. Thus, these data are consistent with the notion that probabilistic splicing provides unique cell identities to L1 and L2 neurons thus preventing inappropriate pairing of postsynaptic elements from the same cell.

As multiple contact synapses are common throughout the fly visual system, we sought to assess whether isoform splicing occurs in a probabilistic fashion more
generally in different classes of visual system neurons. Each of the 12 exon 4 reporters was expressed in a salt and pepper pattern throughout the visual system during periods of synapse formation (Figures 2-6D - 2-6G and Figures 2-S.5B - 2-S.5I). Given the repetitive structure of the visual system (e.g. columns in the lamina and medulla where axons of different cell types branch or terminate in discrete layers; Figure 2-6A), if cell-type specific splicing were a common feature of Dscam1 expression, then we would anticipate observing regular patterns of processes in a columnar or layered arrangement (Figure 2-S.5A). This was not observed, suggesting that splicing is determined neither by cell type nor spatial location in the visual system. Thus, like L1 and L2, it seems likely that the vast majority of visual system cells express alternative variants of exon 4 in a probabilistic fashion. As multiple contact synapses are common throughout the fly visual system, this would be consistent with Dscam1 playing a broad role in regulating their synaptic composition through self-recognition.

Discussion

As differences in Dscam1 expression between neurons is essential for self-avoidance, how isoform expression is regulated is a critical issue in neural circuit assembly in *Drosophila*. The observation that single photoreceptor neurons express multiple isoforms of Dscam1 and that they express different
combinations of them led Chess and co-workers to propose that Dscam1 provided neurons with a unique identity largely through a stochastic process (Neves et al., 2004). These findings were published just prior to our report that Dscam1 isoforms exhibit isoform-specific homophilic binding and followed earlier work from Lee and colleagues describing MB neuron branch segregation defects in Dscam1 mutants (Wang et al., 2002a; Wojtowicz et al., 2004; Wojtowicz et al., 2007). Together these three observations, and our findings that MB neurons also express multiple isoforms and different sets of them (Zhan et al., 2004), led us to propose that sister branches of MB neurons utilize homophilic recognition as a means of discriminating between sister branches and the branches of other neurons. Extensive genetic and biochemical analyses provide a strong case for Dscam1 as a critical determinant of self-avoidance in this and other contexts in the developing fly peripheral and central nervous systems (Hattori et al., 2008; Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007). Ironically, while the nature of Dscam1 expression lies at the heart of self-avoidance, this aspect of Dscam1 biology has remained poorly understood, in large part due to considerable technical challenges.

Here we took a genetic approach to visualize Dscam1 isoform expression. By monitoring exon 4 splicing as a surrogate for isoform expression in neurons, we assessed splicing in vivo broadly throughout the developing nervous system and
in specific cell types, assessed splicing in single identified cells between different animals, and followed splicing in the same cell at different times during development. We demonstrated that splicing is probabilistic in class IV da neurons where it is required for dendritic self-avoidance. The patterns of splicing in both the MB and L1/L2 neurons, where Dscam1 is required for axon branch self-avoidance and appropriate pairing at multiple contact synapses, respectively, are also consistent with a probabilistic mode of splicing.

The analysis of exon 4 splicing in class IV da neurons revealed that, on average each neuron expresses multiple exon 4 variants. The sum of the average expression probability of all the alternative variants of exon 4 in the class IV da neurons is 393 ± 38 % (Figure 2-3K), arguing the splicing of about four variants in a single neuron. If the splicing mechanism were probabilistic, not only at the level of single neurons but also at the level of each round of mRNA processing, it might be expected that the most abundantly spliced variant (i.e. exon 4.2) would be expressed in all neurons, albeit at varying levels, given enough rounds of transcription. Thus, the scattered splicing pattern within a neuronal population may reflect a splicing mechanism in which the same variant is included in multiple mRNAs while others are excluded (e.g. through the assembly of stable splicing complex associated with chromatin). Alternatively, this pattern of
expression may result from low copy numbers of total \textit{Dscam1} mRNAs in each class IV da neuron.

The expression of multiple isoforms in each neuron is a key to robust self-avoidance. Previous studies using RT-PCR analysis on single MB neurons also indicated that each neuron expresses multiple variants of exon 9 (Zhan et al., 2004). Monte Carlo simulations and mathematical modeling suggest that expression of multiple isoforms in a neuron through probabilistic splicing can provide a robust mechanism to endow each neuron with a unique cell surface identity (Forbes et al., 2011; Hattori et al., 2009). Indeed, this robustness is supported by the observation that the differential Dscam1 expression in L1 and L2 neurons arises from probabilistic splicing in these neurons. The large number of isoforms encoded by \textit{Dscam1} is likely to be sufficient to offset the reduced diversity caused by biased exon usage. In addition, dynamic splicing further minimizes the risk that neighboring neurons share the same Dscam1 isoforms for an extended period.

Recent studies suggest that a similar mechanism for self-avoidance has evolved in vertebrates. In the mouse retina and cerebellum, self-avoidance is mediated by a large family of isoform-specific homophilic binding proteins encoded by the clustered protocadherin $\gamma$ locus (Lefebvre et al., 2012). RT-PCR analyses showed
that Purkinje cells express different isoforms, and this is consistent with probabilistic expression of multiple protocadherin \( \gamma \) isoforms in each neuron (Kaneko et al., 2006). Here, regulation appears to be at the level of alternative promoter choice rather than alternative splicing (Tasic et al., 2002; Wang et al., 2002b). Thus, probabilistic expression may have evolved as a common strategy, albeit via different molecular mechanisms, by which neurons acquire unique self-identities.

Although it is clear that Dscam1 plays a prominent role in regulating self-avoidance in multiple contexts, the extraordinary selectivity of homophilic binding and the vast number of different isoforms seems particularly well-suited to mediating recognition between different neurons, for instance between pre and postsynaptic partners. Indeed, Sanes and colleagues have demonstrated that two chick Dscam paralogs (N.B. these genes do not encode multiple isoforms), each with mutually exclusive binding specificities, are expressed in different pairs of synaptic partners in the inner plexiform layers, a structure analogous to the medulla in the fly visual system (Yamagata and Sanes, 2008). Furthermore, gain and loss of function studies support a role for them in matching synaptic partners. If this were the case for Dscam1 isoforms, we would anticipate reproducible cell-type specific patterns of expression of exon 4 variants as homophilic binding requires precise matching of all three variable domains (i.e. encoded by variable
exons 4, 6, and 9). No reproducible patterns of exon 4 expression were observed in any region of the visual system or elsewhere in the developing postembryonic brain. Thus, cell-type specific expression, if it occurs at all, is rare. Thus, it seems unlikely that Dscam1 isoforms selectively regulate matching synaptic partners. Rather, our data provide compelling evidence that a probabilistic mechanism endows cells with unique individual identities throughout the nervous system and is consistent with a widespread role for Dscam1 diversity in patterning neural circuits by preventing inappropriate interactions between axons, dendrites, and postsynaptic elements of the same cell.
Experimental Procedures

**Fly strains:**
Flies carrying membrane-bound GFP tagged with V5 (myristoylated GFP-V5) or nuclear GFP (H2A-GFP) under the control of UAS element were gifts from Barret Pfeiffer and Gerald Rubin (Janelia Farm Research Campus). These were constructed in a vector previously described (Pfeiffer et al., 2010). Class IV da neuron-specific ppk-CD4-tdTom flies (Han et al., 2012) were provided by Yuh Nung Jan (University of California, San Francisco).

**Generation of Dscam1 exon 4 knock-in reporters:**
The knock-in lines were generated in two steps (Figure 2-S.1). First, exons 3-5 were knocked out with a replaceable cassette encoding the white$^+$ marker gene flanked by two attP sites through ends-out homologous recombination, essentially as previously described (Gong and Golic, 2003; Hattori et al., 2007). Homology arms (~3 kb) were amplified from a BAC containing the Dscam1 locus (Venken et al., 2006). In the second step, modified constructs corresponding to the knocked-out region flanked by attB40 were knocked-in through phiC31-mediated RMCE (Bateman et al., 2006; Bateman and Wu, 2008). Alternative variants of exon 4 were mutated by 1bp insertions (Genewiz). The transmembrane domain from mCD8, the T2A peptide (de Felipe et al., 2006), and
Gal4 followed by the SV40 polyA site were fused after exon 5, and the entire construct was cloned in pCR4 (Invitrogen). Injections of the constructs into fly embryos were done at Genetic Services, Inc. and GenetiVision. Successful knock-in events were confirmed by PCR.

**Immunohistochemistry:**

Pupae were staged and dissected as follows. Briefly, white pre-pupae at 0 hr after puparium formation (APF) were collected and incubated at 24°C. Pupal brains were fixed in 4% formaldehyde in PEM buffer (0.1M PIPES, pH6.95, 2mM EGTA, 1mM MgSO₄) for 1 hour at 4°C and washed with PBS-TritonX100 (0.5%), followed by antibody staining in either 10% normal goat serum or 5% normal mouse serum in PBS-TritonX100. Wandering 3rd instar larvae were fixed as previously described (Grueber et al., 2002). Briefly, the animals were filleted along the dorsal midline and fixed in 4% PFA in PBS for 15 min at room temperature. They were washed in PBS-TritonX100 (0.3%). Each step of antibody incubation was performed at 4°C. All stained samples were mounted in EverBrite Mounting Medium (Biotium) and imaged with Carl Zeiss LSM 510 Meta. Microscope settings for the GFP channel were kept constant for all samples requiring comparison. Antibodies used were as follows: Alexa Fluor 488 conjugated mouse anti-V5 at 1:200 (Serotec); mouse anti-Dachshund at 1:20 (Developmental Studies Hybridoma Bank (DSHB)); mouse anti-Fasciclin II at
1:20 (DSHB); rat anti-Ncadherin at 1:20 (DSHB); mouse anti-Seven-up at 1:10 (a gift from Yasushi Hiromi, NIG, Japan); chick anti-GFP at 1:1,000 (Invitrogen); rabbit anti-DsRed at 1:200 (Clontech); Alexa Fluor 647 goat anti-mouse at 1:200 (Invitrogen); Alexa Fluor 568 goat anti-rat at 1:200 (Invitrogen); Alexa Fluor 568 goat anti-rabbit at 1:200 (Invitrogen); and Alexa Fluor 488 goat anti-chick at 1:200 (Invitrogen).

**Live imaging:**

Live imaging of the class IV da neurons was done using a custom built 2-photon microscope. Larvae mounted in a mix of Halocarbon Oil 27 and 700 (1:1) were anesthetized with CO2, and dorsal rows of class IV da neurons were identified using tdTom. The first imaging session was done at 48 hours after hatching, after which the animals were put back into 24°C incubator in separate chambers. The second imaging session was done at ~96 hours after hatching.

**Image analysis:**

For the analysis of mushroom body, z-stacks were acquired with a 40X objective such that the entire depth of the lobes was captured as assessed by Fasciclin II staining, and maximum intensity projections were obtained. For the visual system, z-stacks of the lamina and the medulla from dorsal to ventral were acquired with 63X and 40X objectives, respectively, at similar locations in all brains. Either 5
µm (lamina) or 10 µm (medulla) maximum intensity projections for similar locations were generated using Fiji.

For the analysis of stained class IV da neurons, 7-10 animals per genotype were analyzed. CD4-tdTm was used to detect the location of the nucleus and to determine a single confocal plane where the diameter was at maximum. A circular region of interest with a set diameter was placed over each nucleus, and the average GFP intensity within the region was quantified using Fiji. Subsequent statistical analyses were done in R. In stained samples, images were acquired using 40X objective with 6X zoom, and the circular region of interest was set at 23.716 µm². Background GFP intensity was also measured in the cytosolic region and was subtracted from the nuclear GFP intensity to compensate for sample variation. The threshold between ON and OFF was determined such that most GFP values from eight negative control animals were OFF; when negative control values are fit with Gamma probability density function (PDF), the threshold corresponds to p-value of 0.0035. Animals for exons 4.4, 4.5, and 4.11 reporter were prepared separately from the other exons, together with 10 animals each for exons 4.1 and 4.6 as references. Threshold was adjusted such that both distributions for exons 4.1 and 4.6 were statistically not different between the different batches.
For live imaging, a stricter threshold was used. First, GFP values from negative control values were fit with gamma PDF, and the OFF threshold value was determined at p-value of 0.1. All values below this threshold were considered OFF. For a cell to be considered ON, the GFP value had to be greater than another threshold drawn at p-value of 0.005. All values between p-value of 0.1 and 0.005 were discarded from the analysis as ambiguous values.
Supplementary Information

Figure 2-S.1 | Generation of Splicing Reporter Knock-ins by Homologous Recombination Followed by phiC31 Recombination Mediated Cassette Exchange

Using ends-out homologous recombination, endogenous Dscam1 exons 3-5 were replaced with DNA encoding white\(^+\) (a marker gene used for screening) that is flanked with inverted phiC31 recombination recipient sites (attP). In the second step, genomic sequences flanked with recombination donor sites (attB40) containing exons 3-5 with various modifications to exon 4 and exon 5 were inserted back into the original site by phiC31 recombinase, thereby replacing white\(^+\). The white\(^+\) insertion allele is a null for Dscam1 and recessive lethal. Replacement with a wild type cassette rescues the phenotype, as expected. This indicates that the modifications in the introns resulting from the attR sites do not disrupt Dscam1 function. This supports the view that the reporters accurately reflect expression.
**Figure 2-S.2 | Expression of Alternative Variants of Exon 4 Are Variable in the MB**

Expression of the splicing reporters for each alternative exon 4 variant in the MB. All 12 splicing reporters were examined. Data for four of these are shown in the main figures (Figure 2-2). The remaining eight are shown in this figure, together with exon 4.1 as a reference. The samples shown in panels (D-J) were prepared separately from the other samples shown in (A-C) and the main figure (Figure 2-2), and thus, exon 4.1 was included in both experiments as a reference. White, V5-tagged membrane-bound GFP driven by each splicing reporter; blue, mushroom body neurites (visualized by anti-Fasciclin II) or Kenyon cell nuclei (visualized by anti-Dachshund). Scale bars: 40 µm, MB lobes; 30 µm, Kenyon cell bodies.
H

<table>
<thead>
<tr>
<th></th>
<th>Negative</th>
<th></th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Exon 4.1</td>
<td>*</td>
<td>*</td>
<td>Exon 4.1</td>
</tr>
<tr>
<td>Exon 4.2</td>
<td>*</td>
<td>*</td>
<td>Exon 4.2</td>
</tr>
<tr>
<td>Exon 4.3</td>
<td>*</td>
<td>*</td>
<td>Exon 4.3</td>
</tr>
<tr>
<td>Exon 4.4</td>
<td>*</td>
<td>*</td>
<td>Exon 4.4</td>
</tr>
<tr>
<td>Exon 4.5</td>
<td>*</td>
<td>*</td>
<td>Exon 4.5</td>
</tr>
<tr>
<td>Exon 4.6</td>
<td>*</td>
<td>*</td>
<td>Exon 4.6</td>
</tr>
<tr>
<td>Exon 4.7</td>
<td>*</td>
<td>*</td>
<td>Exon 4.7</td>
</tr>
<tr>
<td>Exon 4.8</td>
<td>*</td>
<td>*</td>
<td>Exon 4.8</td>
</tr>
<tr>
<td>Exon 4.9</td>
<td>*</td>
<td>*</td>
<td>Exon 4.9</td>
</tr>
<tr>
<td>Exon 4.10</td>
<td>*</td>
<td>*</td>
<td>Exon 4.10</td>
</tr>
<tr>
<td>Exon 4.11</td>
<td>*</td>
<td>*</td>
<td>Exon 4.11</td>
</tr>
<tr>
<td>Exon 4.12</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2-S.3 | Each Alternative Variant of Exon 4 Is Expressed in a Subset of Class IV da Neurons in a Probabilistic Fashion

(A-G) Expression of all 12 variants of exon 4 were analyzed in the class IV da neurons. Five are shown in the main figure (Figure 2-3), while the other seven are shown here. The frequency of cells expressing different alternative variants differs. Blue indicates ON, red indicates OFF, and grey indicates neurons that could not be scored.

(H) The statistical significance between expression frequencies per animal for each pair of alternative variants of exon 4 is shown as calculated for Figure 2-3K. Kruskal-Wallis followed by Steel-Dwass. * indicates p < 0.05.

(I) Expression of the controls, exon 4.2, and exon 4.6 in overlapping class III and class IV neurons in the ventral cluster. Each rectangle (representing the ventral cluster) is divided into two triangles with the lower left triangle indicating the expression in class III neurons, and the upper right triangle indicating the expression in class IV neurons. No correlation was observed between the expression in the two neurons.
A  Experimental data set  Randomized data set

Shuffle X 1,000  Distribution of 1,000 PCC averages

Average pairwise correlation

B

Exon 4.3  Exon 4.4  Exon 4.5  Exon 4.6

Exon 4.7  Exon 4.8  Exon 4.10  Exon 4.11

p = 0.225  p = 0.029  p = 0.033  p = 0.370

p = 0.267  p = 0.434  p = 0.322  p = 0.093
Figure 2-S.4 | No Significant Bias in the Spatial Expression Pattern of Each Alternative Variant of Exon 4 in the Class IV da Neurons Was Detected

(A) Workflow of the statistical test. Experimental data set for each alternative exon was shuffled to generate a randomized set, and the average of pairwise Pearson correlation coefficients (PCC) within the set was calculated. This process was repeated one thousand times to generate a distribution of the average PCC for the randomized sets. This distribution was then fit with a normal distribution to be compared to the average PCC from the experimental data set.

(B) Comparison of the average PCC within the experimental data set (red line) with the distribution of average PCC calculated for 1,000 independent sets of randomized data in blue. None of the differences between the experimental and the randomized data set were statistically significant (p>0.025; two-tailed). Results for exons 4.1, 4.2, 4.9, and 4.12 are shown in the main figures (Figure 2-4). P-values are given on the top right hand corner.
Figure 2-S.5 | Scattered Subsets of Specific Cell Types Express Alternative Variants of Exon 4 in the Visual System

(A) An example of a cell-type specific marker that labels all L3 neurons (green). As L3 neuron is repeated in every column, and as its axon terminates in a specific layer in the medulla, the marker results in a uniform layer across the medulla. Layered patterns similar to this are seen with many other cell-type specific markers. Uniform columnar patterns are also seen with the photoreceptor axons (blue). Such patterns were never observed with the splicing reporters for exon 4.

(B-I) Expression of each splicing reporter in the lamina and the medulla of the fly visual system. Exons 4.1, 4.2, 4.9, and 4.12 are shown in the main figures (Figures 2-6B - 2-6G), and all the others are shown here. All alternative variants are observed in the L1 neurons (yellow arrowheads) but are expressed only in a subset of them. Neither uniformly layered nor repeated columnar patterns were apparent in the medulla (B_{2-I_{2}}). Thus, expression of a specific exon 4 variant in all cells of a specific type in the visual system occurs rarely or not at all. White, membrane-bound GFP tagged with V5 epitopes; cyan, L1 nuclei (anti-seven-up); blue, neuropile (anti-N-cadherin). Scale bars: 15 µm, lamina (B_{1-I_{1}}); 30 µm, medulla (B_{2-I_{2}}).
Appendix 2-1: Generation of the Recipient Fly for phiC31-RMCE Through Ends-Out Homologous Recombination

The splicing reporters for the alternative splicing of Dscam1 exon 4 cluster were knocked into the endogenous locus through phiC31-RMCE with the recipient fly, which had the region spanning exons 3-5 replaced with a replaceable cassette encoding white+ (Bateman et al., 2006). This recipient fly was generated through ends-out homologous recombination, following a similar protocol described by Rong and Golic (Figure 2-A.1A) (Rong and Golic, 2000). Briefly, 3kb homologous arms were cloned from a BAC containing Dscam1 and attached to the ends of the replaceable cassette, which consisted of inverted attP sites flanking mini-white+ (Venken et al., 2006). Based on the genomic sequence comparison among various Drosophila species, regions with the least conservation were chosen as the insertion sites for the attP sequences. The construct was then cloned into a P-element vector, in between two direct-repeat FRT sequences. After insertion of the vector into the genome as a P-element, the construct was flipped out in the germline with heat shock inducible FLP and linearized with a restriction enzyme Isce-I to expose the two ends of the homologous arms. The resulting flies were screened for gain of white+ phenotype (note that the flies become white+ once the construct is flipped out, unless it is inserted back into the genome).
In the initial trial, 14 potential recombinant flies were recovered. However, genomic PCR revealed that in all of the lines, only one of the two homologous arms had recombined with the correct region in the genome, whereas the other arm was inserted into unidentified regions. It was essentially random as to which arm had recombined correctly. Inverse PCR showed that the non-recombined arm had a vector sequence still attached at the end, indicating that Isce-I digestion failed to expose the homologous end (Figure 2-A.1A).

We reasoned that the failure to recombine both ends resulted from inefficient digestion by Isce-I. To expose both ends of the homologous arms, two digestion events were required in the initial design. Thus, in the second trial, the original scheme was revised such that only one Isce-I cut was required to expose the two ends (Figure 2-A.1B). Briefly, this was achieved by joining the two homologous arms together with Isce-I site in the middle, such that 3’ arm continued into the 5’ arm. The whole piece was then cloned into one side of the mini-white^+ cassette. Once flipped out through FLP mediated FRT recombination in the germline, this results in a circular DNA that requires just one Isce-I cut to reveal the two ends of the homologous arms. This experiment resulted in 8 potential candidates, and though the issue of one-sided recombination still persisted, we were able to recover 2 recombinants with both arms recombined at the correct site.
Figure 2-A.1 Strategies of Ends-Out Homologous Recombination

(A) Conventional ends-out homologous recombination. In the donor construct, two homologous arms were placed on either side of the replaceable cassette encoding *mini-white*. After flipping out the donor in the germline via FRT recombination, the ends of the homologous arms are joined by a piece of DNA encoding FRT. This requires two Isce-I digestion events to expose the ends of the arms. All potential recombinants obtained from this experiment had only one of the homologous arms recombined with the genomic locus. The other arm was inserted through non-homologous end joining. Inverse PCR showed that the FRT sequence was still attached to the end of the arm that did not recombine (red box).

(B) To circumvent the problem of inefficient Isce-I digestion, the layout of the homologous arms in the donor construct was changed. The two homologous arms were joined together with Isce-I site, with the 3’ homologous arm located upstream of the 5’ homologous arm. The attP sites used to make the *mini-white* cassette replaceable were located on either side of the joined arms. After the flip-out event, this construct becomes a circular DNA that requires just one Isce-I cut to expose the ends of the arms.
Replacement of the cassette with a wildtype genomic sequence rescues the lethality of the locus.

The revised scheme also offers an additional benefit. The construct, after linearization, harbors an FRT site next to the replaceable cassette. Thus, in theory, this enables the generation of desired recombinants through FRT mediated recombination in the germline between two one-sided recombinants, one whose 5’ arm has correctly recombined and the other, 3’ arm (Figure 2-A.1C).
References


Chapter 3

Discussion
Probabilistic Splicing of Dscam1 Is a Robust Means to Achieve Unique Identity at the Level of Single Neurons

Self-avoidance was first demonstrated by Kramer and his colleagues in the mechanosensory neurons of the peripheral nervous system of the leech, *Haementeria ghilianii* (Kramer and Kuwada, 1983; Kramer and Stent, 1985). The studies were conducted on a neuron that sends out three parallel primary branches, each giving rise to higher order branches to form three adjacent receptive fields that exhibit little overlap and sharp boundaries. This mutual exclusion of the receptive fields, and the observation that the sister processes within a receptive field also do not cross each other, suggested that the processes belonging to the same neuron had the tendency to repel one another. Indeed, ablation of one of the receptive fields early in development lead to the invasion of the vacated region by the adjacent sister receptive field. Importantly, processes of the neuron freely overlapped with those of neighboring homologous neurons, indicating that the repulsive interaction was strictly between the processes of the same neuron. To explain this phenomenon, Kramer and his colleagues proposed that processes carried “cell-specific, or idiosyncratic, labels” that allowed them to distinguish self from non-self. They noted that “the number of such labels would have to be quite large” and that each cell “would have to carry a different set of labels that would allow its own isoneuronal recognition to
occur independently of any heteroneuronal interactions” (Kramer and Stent, 1985). However, due to the enormous complexity that such labels would require, an alternative discrimination mechanism based on differences in neuronal activity patterns was favored at the time.

Mounting evidence in the last decade has established Dscam1 as the mediator of self-avoidance in Drosophila. Dscam1 is capable of generating an enormous degree of diversity through alternative splicing, and each isoform exhibits isoform-specific homophilic binding with very little heterophilic affinity. Thus, through combinatorial expression of a small subset of the isoforms in each neuron, the gene is capable of generating a large number of unique cell surface identities. This allows the processes of a neuron to discriminate between those that originated from the same cell body (“self”) and those that originated from different neurons (“non-self”). Our analyses using splicing reporters for each alternative variant demonstrate that this acquisition of unique identities is achieved through probabilistic splicing. Thus, the enormous complexity of “cell-specific, or idiosyncratic, labels” is in fact solved in a conceptually simple fashion.

Assuming probabilistic splicing for all three exon clusters corresponding to the variable Ig domains, how many unique identities can be supported by Dscam1? Previous Monte-Carlo simulations and mathematical solutions provide some
estimation to this problem. First, under the assumptions that 1) each neuron makes 30 isoforms and 2) two neurons must not share any isoform, only a few neurons can be supported even with 20,000 isoforms, as the probability that a pair of neurons express the same isoform is counter-intuitively high (Hattori et al., 2009). Thus, this argues that some degree of overlap in two neighboring neurons must be permitted. Both the simulations by Hattori et al. and a mathematical solution presented by Forbes et al. argue that more than ten thousand distinct identities can be supported by 10,000 isoforms and tens of thousands by 20,000 isoforms when 20% overlap is allowed (Forbes et al., 2011). It should be noted that these studies make a simplifying assumption that splicing is purely random (i.e., that there is no bias). The bias could in effect decrease the number of unique combinations, and thus, the number may be an overestimate. Nonetheless, the estimations are consistent with the observation that thousands of isoforms are required to support unique self-identities in the MB that consists of some 2,500 neurons.

Probabilistic splicing of Dscam1 is seen throughout the nervous system. Self-avoidance is required for many aspects of the circuit assembly, such as regulating synaptic composition at multiple-contact synapses in the visual system and promoting uniform coverage of dendritic fields in the sensory neurons (Hughes et al., 2007; Matthews et al., 2007; Millard et al., 2010; Soba et al.,
2007). In *Drosophila*, multiple-contact synapses, with two or more postsynaptic elements opposing a single presynaptic release site, are prevalent throughout the central nervous system and are the main type of synaptic organization (Prokop and Meinertzhagen, 2006). Thus, Dscam1 mediated self-recognition may play a widespread role in establishing the appropriate composition of synapses during the assembly of the neural circuits.

Probabilistic expression of Dscam1 isoforms in each neuron means that the isoforms are unlikely to mediate interactions between the processes of different neurons, as such interactions would require a prior knowledge of the isoforms the interacting partner has chosen to express. Thus, it seems that each neuron indeed acquires “individual identification tags… to the level of the single neuron.” Yet, unlike Sperry envisioned decades ago, the main function of these individual tags seem to be mainly for mediating interactions between the processes of the same cell, and not between those of different cells.

**Unresolved Issues on Dscam1 Splicing**

Our data on the splicing of Dscam1 leads us to a couple of interesting issues to be addressed in the future: What is the molecular mechanism behind the
probabilistic splicing, and is there a functional significance to the different splicing biases observed in different tissues?

The molecular mechanisms by which probabilistic expression of \textit{Dscam1} arises remain largely unexplored. Based on the genomic sequence comparisons among 16 different insect species that are separated by \textasciitilde300 million years, Graveley has proposed a model for the mutually exclusive splicing from the exon 6 cluster (Graveley, 2005). This involves a selector sequence located adjacent to each variant and a single docking site of a complementary sequence. A variant is selected based on the docking site:selector sequence pairing, and this mutually exclusive pairing ensures that only one variant is included in the mature mRNA. Similar mechanisms have also been proposed for exon clusters 4 and 9 (Yang et al., 2011). Thus, probabilistic splicing could arise from probabilistic pairing of the docking site with one of the selector sequences. Alternatively, it could arise from probabilistic expression of the splicing factors that determine the variant choice. RNAi screens of RNA binding factors in \textit{Drosophila} cultured cells have lead to the identification of several candidate factors that affect the inclusion frequency of different variants (Park et al., 2004). How they contribute to the splicing of \textit{Dscam1} mechanistically and whether this contribution holds any significance \textit{in vivo} are only beginning to be explored. These questions could be addressed in the future with a combination of biochemistry and the splicing reporters. For
instance, the reporters could be used to monitor the effects of various perturbations, including deletions of regulatory sequences for splicing and mutations in the splicing factor genes.

Tissue-specific bias in variant choice has been observed in multiple contexts (Celotto and Graveley, 2001; Neves et al., 2004; Zhan et al., 2004). Microarray studies have detected differential bias of exon 9 splicing in the photoreceptor cells and the MB neurons. In the current study, we have also identified a strong bias in exon 4 splicing among different neuronal tissues. A striking difference in exon 4.2 splicing can be observed between the MB and the da neurons. Furthermore, two specific classes of glial cells in the visual system present an extreme case where only exon 4.2 but no other variant is spliced. Do these biases hold any functional significance? Specific null mutation of exon 4.2 did not produce any gross morphological phenotype in the glial cells at the level of light microscopy (S.K.M. and S.L.Z., unpublished observation). On the other hand, from analyses of the alternative exon 4 splicing in tissue preparations (i.e., antennae, heads, wings, and legs) of *Drosophila yakuba*, separated from *melanogaster* by about 10 million years, Celotto and Graveley concluded that the bias is evolutionarily conserved, potentially indicating a functional significance (Celotto and Graveley, 2001). One prevailing hypothesis is that conferring different bias in different neurons is important to minimize the probability that the
two neurons share the same Dscam1 isoforms. This may be especially relevant for a pair of pre- and postsynaptic neurons. Thus, it is possible that more detailed analyses, such as of the cellular compositions at multiple-contact synapses at the level of electron microscopy, are required to uncover the roles of the splicing bias. Identification of the splicing factors in the future may allow us to alter the bias in specific cell types to test the functional significance.

**Evolution of Self-Avoidance and the Responsible Genes**

Recent studies by Sanes and his colleagues have identified clustered protocadherin $\gamma$ as the mediator of self-avoidance in the vertebrates. Though evolutionarily unrelated to *Dscam1*, protocadherin $\gamma$ exhibits some of the main features critical for self-avoidance. First, they encode multiple isoforms from each locus through alternative selection of the first exon that is joined to common 3’ exons (Tasic et al., 2002; Wang et al., 2002). This allows the gene to produce 22 isoforms from a single locus. Second, biochemical evidence suggests that these isoforms form cis-multimers that bind homophilically to other multimers of the same composition, thus generating a large number of distinct combinations (Schreiner and Weiner, 2010). Furthermore, RT-PCR analyses in the Purkinje cells are consistent with probabilistic expression of multiple protocadherin $\gamma$ isoforms in each neuron (Kaneko et al., 2006). Knock-out of protocadherin $\gamma$ in
the starburst amacrine cells in the retina and Purkinje cells produce phenotypes highly similar to those observed in Dscam1 null neurons in Drosophila, exhibiting defects in self-avoidance. Thus, the vertebrates seem to have evolved an independent yet very similar mechanism to Dscam1 to achieve self-avoidance.

The clustered protocadherins, which possess six cadherin repeats, have not been found outside of the vertebrates, though related non-clustered protocadherins with seven cadherin repeats have been found in the protostomes and even in a cnidarian (Hulpiau and van Roy, 2011). On the other hand, the extensive diversification of Dscam1 gene seen in insects and crustaceans has not been identified outside Pancrustacea, though its homologs without the isoform diversity are present throughout the deuterostomes. Thus, the diversification of Dscam1 seems to be a Pancrustacea-specific innovation that emerged after the split from other arthropods (Armitage et al., 2012). It is interesting that self-avoidance was initially described in the leech, yet its genome seems to encode neither the clustered protocadherins nor the diversified Dscam1 isoforms. Thus, discrimination of self and non-self must be achieved through a yet alternative means in the leech. Furthermore, the evolutionary tree predicts that several other groups with complex nervous systems, such as mollusks, also possess neither of the diverse molecular families (i.e., Dscam1 and Protocadherins). This suggests that through the course of evolution, multiple
mechanisms evolved independently to mediate self-avoidance in different groups. Thus, the mechanism to discriminate self from non-self may not have existed in the last common ancestor of the bilaterians, and it may be a more recent invention that coincides with the growing complexity of the nervous system.
References


