

SEPTEMBER 21, 2007

## Gene Assigns ID Tags to Help Organize the Developing Brain

The developing nervous system is a seemingly chaotic and exceedingly complex jumble of cells with specialized missions, unique architectures, and stereotyped patterns of neuronal connections, or synapses. How neurons' dendrites and axons weave themselves into precise neural circuits during development remains a challenging question in neurobiology. What are the molecular tags on the surface of neurons that allow them to distinguish between each other?

A single gene capable of producing more than 38,000 cell surface proteins is an essential tool in assuring the assembly of precise neural circuits in the fruit fly, *Drosophila melanogaster*. Now, two teams of researchers from the Howard Hughes Medical Institute (HHMI) have demonstrated how these closely related proteins establish the specificity that allows them to serve as identification tags for individual neurons.

---

"Expanding from 11 to 19,000 is a bit of a stretch. So the question was, how many different binding specificities are there?"

— S. Lawrence Zipursky

---

In work published in the September 21, 2007, issue of the journal *Cell*, research teams led by HHMI researchers S. Lawrence Zipursky of the University of California at Los Angeles (UCLA) and David Baker of the University of Washington worked together to describe how each of 18,048 different versions of the Dscam protein is able to recognize and bind only to an identical form of the protein.

Dscam — short for Down Syndrome Cell Adhesion Molecule — was discovered in 2000 in a collaborative effort by the Zipursky lab and the lab of Jack Dixon, who was then at the University of Michigan and is now Vice President and Chief Scientific Officer at HHMI. Work in Zipursky's lab and others has shown that Dscam is essential for the establishment of neural circuits throughout the fly nervous system. Dscam allows the dendrites, arbor-like structures that sprout from nerve cells, to distinguish between other arbors from the same cell and arbors from neighboring cells; arbors from the same cell then selectively grow away from each other. This process is known

as self avoidance.

In 2004, Zipursky and his colleagues proposed that Dscam did so because each of 19,008 subtly different versions of the protein was able to recognize and bind only to a Dscam molecule identical to itself. With each neuron expressing a different, random set of Dscam molecules on its surface, it could then differentiate between its neighbors and itself.

This idea was supported by genetic studies and biochemical tests of 11 different forms of Dscam, each of which would bind to itself, but not to any of the other 10 versions. But, Zipursky said, Expanding from 11 to 19,000 is a bit of a stretch. So the question was, how many different binding specificities are there? And, on a molecular, biochemical, and atomic level, how can that specificity be achieved?

We couldn't test 19,000 Dscam isoforms, Zipursky said, so we wanted to come up with a way we could test a very large number and make a reasonable estimate as to how many binding specificities there would be.

The modular construction of Dscam proteins helped simplify the task. The part of the protein that sits outside of the cell and interacts with other Dscam proteins includes three variable modules, known as Ig2, Ig3, and Ig7, Zipursky explained. As a cell assembles a Dscam protein, it can choose from 12 different versions of Ig2, 48 versions of Ig3, and 33 versions of Ig7. All possible combinations of these three domains yield 19,008 slightly different forms of Dscam. Two different forms of the part of the protein that anchors Dscam to the surface of the neuron further increase Dscam variability, but this part of the protein does not affect binding specificity.

Zipursky and his colleagues showed that when two Dscam proteins bind, the Ig2 domains match up to one another. Likewise, one protein's Ig3 domain binds to the Ig3 domain on the other, and the Ig7 domains do the same. So the researchers analyzed the binding specificity of each of these domains individually. They tested all of the forms of Ig2 against one another, and then did the same for Ig3 and Ig7. It's a huge grid - thousands of interactions. Zipursky said. But it's not 19,008 times 19,008.

The researchers found that each Ig2 could bind only to an identical Ig2. Their results were similar when they tested the 48 versions of Ig3 and the 33 versions of Ig7. About five percent of the non-identical pairs showed binding, but it was always weaker than the binding of each form to itself. From this, they inferred that 95 percent of the 19,008 Dscam forms had the same highly specific self-binding characteristic they had found in their first set of 11 proteins. This enormous resource of molecular recognition can then be used to wire the brain, Zipursky said.

So far, Zipursky and others have demonstrated that neurons use Dscam to recognize and avoid sister dendrites (dendrites from the same cell) as well as sister arbors of axons. He says it's also possible that in some developmental contexts neurons use Dscam proteins to specifically recognize other neurons.

So how do you get this kind self-binding specificity for such a huge group of proteins? Zipursky asked. If it were five or ten, you'd say okay, that's interesting. But when it's in the tens of thousands, it's really quite mind-boggling.

To begin to answer that question, Zipursky's group began with recent work from a group at Harvard led by Jia-huai Wang and Dietmar Schmucker, who had used x-ray crystallography to reveal part of the molecular structure of one pair of Dscam proteins bound to one another. The structure showed how short segments of Ig2 and Ig3 domains on one Dscam protein bound to the same segments of the partner molecule, with the segments in the two proteins flipped in opposite orientations. If you put your arms in front of you and cross them so that your elbow is touching the fist of the other arm, and so too for the other, Zipursky explained, that's the way they interact with each other. They have to line up in a very precise anti-parallel way in order for them to bind.

To test whether these segments were the key to Dscam's binding specificity, Zipursky and his colleagues plucked them out of the proteins and swapped them for one another. This was done for many Ig2 and Ig3 domains. Remarkably, this was enough to swap the binding specificity. The scientists concluded that each form uses a similar strategy for matching as seen in the X-ray structure of the one form solved by Wang and Schmucker. The specific properties of these short sequences, which are different between the 12 alternative Ig2 and 48 alternative Ig3 domains, determine the specificity of binding.

Since no structural data were available for the Ig7 domain, the group compared that part of Dscam to related molecules for which X-ray structures were known. By comparing the sequences of the few closely related forms that showed weak binding, they predicted which portion might be critical for binding. They tested their predictions by making targeted changes in the proteins' sequences, and found that, unlike in Ig2 or Ig3, critical binding sequences appeared on two separate strands of the protein. Making these changes between isoforms swapped the binding specificity.

To learn more about how Ig7 might work, they turned to HHMI investigator David Baker, an expert in protein modeling at the University of Washington who had already expressed a fascination with Dscam. When he first heard of the remarkable binding specificity of Dscam's many forms, Baker said, I couldn't imagine how that could possibly be the case. It sounded like science fiction to me.

We have no structure for Ig7, Zipursky said he told Baker. We have some biochemical data, and we'd love to have you work out how these might interact with each other. That's a real challenge and not something we were even remotely capable of doing. Through protein modeling, Baker found that Dscam's Ig7 domains interact with one another using the same antiparallel alignment as the protein's other two variable domains, but use a different structural strategy to do so.

The ability of Dscam to specify and identify so many cells so precisely is a wonder of nature, according to Baker. "It demonstrates biochemically that Dscam variants have a huge number of recognition specificities. It's quite unparalleled in biology," he explained.

Indeed, with the exception of cells in the vertebrate immune system, Dscam seems to power a more robust cell recognition and identification system than has been reported in any other cell type, Zipursky said.

The new work was conducted using proteins from the fruit fly model *Drosophila melanogaster*. While tens of thousands of Dscam isoforms can be made in all insects whose genomes have been sequenced so far, mammalian genomes only encode a few. But it seems quite likely, according to Baker and Zipursky, that other large protein families may have evolved similar functions as molecular identification tags in the developing central nervous systems of other animals, including humans.

And while Dscam seems to be a key player in guiding the formation of precise neural circuits in *Drosophila* it is not the whole story, Zipursky noted: There are other genes and proteins at work that allow neurons to selectively recognize each other. Identifying these proteins and understanding how they all work together is an enormous challenge.

In addition to Baker and Zipursky, authors of the *Cell* paper include Woj M. Wojtowicz and Wei Wu of UCLA, and Ingemar Andre and Bin Qian of the University of Washington.