Memory: From Sensory Circuits To Protein Conformations

by

John P. McGinnis

Submitted to the graduate degree program in Molecular and Integrative Physiology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____

Co-Chair Kausik Si

____

Co-Chair Gustavo Blanco

____

Aron Fenton

____

Timothy Fields

____

Robb Krumlauf

____

Ting Xie

____

C. Ron Yu

Date Defended: January 9, 2017

The Dissertation Committee for John P. McGinnis

certifies that this is the approved version of the following dissertation:

Memory: From Sensory Circuits To Protein Conformations

Co-Chair Kausik Si

Co-Chair Gustavo Blanco

Date approved: 04/10/2017

Abstract

The ability to form, store, and retrieve memories is an essential capacity of many animals. Only in the past half-century, however, have the key neuronal and molecular events that underlie memory been studied. This has revealed a series of molecular cascades that are triggered by neuronal activation, ultimately leading to the stabilization of otherwise transient synaptic modifications. These synaptic modifications either increase or decrease the efficiency of synaptic transmission, thereby leading to altered neuronal communication. The behavioral events that must precede these molecular and neuronal changes, however, begin with the sensory system, where important information from the external world is identified. Generally, in situations involving associative learning, the experience of a reward or punishment is assumed to be remembered because of the relevance it holds for the organism at the moment of learning. Any pattern that deviates from this general idea provides an indication of exactly which features the processes of memory deem valuable. The deviation we explore here—that a more immediately appealing reward is not always better remembered, that D-arabinose is preferred to L-arabinose, but L-arabinose generates more reliable memories—suggests that there are important aspects beyond the momentary appeal of a reward. Further, L-arabinose, because it is remembered despite having no nutritional value, has allowed identification of a subset of 26 sensory neurons in *Drosophila* that, when activated, are sufficient to form long-term associative memories. In response to this sensory activation, the biochemical cascades of memory will, in further downstream neurons, trigger the oligomerization of the *Drosophila* cytoplasmic polyadenylation element binding (CPEB) protein, Orb2. Its oligomerization has so far been described as involving a prion-like conversion from a monomeric form to an amyloidogenic oligomer. Since its initial characterization, however, many types of functional protein oligomerization have been described. The question of whether Orb2's oligomerization is in fact prion-like can best be addressed by substituting its prion domain with a variety of oligomerization domains, from other amyloid domains to more transient 'liquid droplet'-type domains, or even standard tetramerization domains. Whether formation of less structured and less stable aggregates can still support the regulatory

switch of Orb2 required for memory maintenance is the key question. Finally, Orb2 may not be the only functional prion-like protein in *Drosophila*, though it is the best characterized. Murashka, a RING-domain E3 ubiquitin ligase, is also involved in memory, and possesses both a disordered domain and features that are characteristic of prion-like proteins. If murashka does in fact undergo a prion-like conversion that is relevant for its role in memory, Orb2 will no longer be a curious outlier but instead the first illustration of what may be a widespread biological phenomenon.

Acknowledgments

I owe a great deal to a great many people, and consider myself absurdly lucky in a huge number of ways. I don't think you can ever actually deserve what luck has given to you, but I'll say that I try to work hard not to let it go completely to waste.

My interest in biology really began with Paul Wieland, my AP Biology (and Anatomy and Physiology) teacher. It was he and those classes that made me think I could do something with biology as a career. Luther College's biology faculty was full of fantastic teachers, especially Scott Carlson and Marian Kaehler. Carlson and David Bishop were the two professors who were willing to involve me in their respective research programs—my first research experiences. In both cases, the work was fascinating: surgery on rats for one, and a research poster presented in Kyoto, Japan for the other. I owe them both for any career I'll have in research.

Tim Fields, the MD-PhD program director—without him I would not be writing this. I still remember exactly where in Luther's science building I was standing when he called, asking if I was interested in interviewing. Saying yes was one of the best decisions of my life. He treated Emily and I like we had every option in the world, even though we didn't. Once I was in the program, he again intervened when I needed to find a lab: he kept insisting to Kausik—who was otherwise not going to take me, I learned after I defended—that Kausik should really give me a chance. For these two times alone, I owe him everything.

Without current and former Si lab members, I would have floundered. Amitabha Majumdar patiently answered every fly question I thought of; Huoqing Jiang and Wanda Colon-Cesario took time to train me in fly behavior assays; Ed Choi was a fun and very helpful baymate. Jason Gill was the first one to introduce me to molecular biology, and I learned my first months in lab by helping him with his projects; he was a fantastic and thankfully patient mentor. Liying Li has been the best graduate student in the lab my entire time there, and easily the most helpful. No matter the topic—fly genetics, behavior, molecular

biology—any time I have a question Liying is the first person I ask. And the most amazing result I ever got in lab—turning on taste neurons and generating memory—was completely thanks to her. We had tried for two years, but it wasn't until she read a paper detailing a new channelrhodopsin and showed me that we got it to work. Gunther Hollopeter wasn't technically a Si lab member, but it always felt like he was. Thanks to him for the huge number of discussions about science, careers, and how to think about scientific problems. Danny and Angie Miller are, first, great friends, despite how strenuously Danny tries to deny it. But thanks to him for what little I know about genomics, which is just enough to keep pestering him with more questions. I suspect I'll be grateful over the next fifty years for having even a shallow familiarity with the area. And finally to our collaborators in Paris: Frederic Marion-Poll, Moutaz Ali Agha, and Palak Rawat—with their assistance, our paper was much stronger.

And of course my family, Ellen, Tim, Jenny, and Katie. It wasn't until years later that I realized how lucky I am to be a part of this family. Here's to many more great memories.

To Kausik I owe the world. I have no idea where I would be if he hadn't (finally) agreed to take me. I had to scramble to find a lab at what seemed like the last minute, and though I had had my eye on Kausik's lab earlier than that, I still can't believe my luck at ending up in his lab. The questions, the techniques, the style of thought, the atmosphere of the lab—everything has fit so well. I enjoyed my time in lab far, far more than I ever expected to. But the best thing about joining the lab was getting to know Kausik and learning from him. We have gotten along so well, and trying to stretch my mind to understand how he thinks and how he approaches problems has made me so much smarter than I ever would have been otherwise.

Finally, Emily and now Ella. Nothing I write can say how grateful and happy I am and how lucky I feel to have you two.

TABLE OF CONTENTS

GENERAL INTRODUCTION 1
The cellular and molecular basis of memory 3
Drosophila's role in memory research 5
Where are memories located? 7

PART I. REWARD PERCEPTION AND LONG-TERM MEMORY 10
Reward components 11
Food as a rewarding memory 12
Chemosensory perception in Drosophila 13

RESULTS 15
Flies rapidly evaluate the value of a sugar 15
Flies prefer D-arabinose but form more reliable memories of L-arabinose 16
The behavioral difference between D- and L-arabinose is in the chemical nature of the sugars 18
The neural basis for the difference in behavioral responses to D- and L-arabinose 26
Peripheral Gr43a neurons are necessary for L-arabinose memory 34
L-arabinose and D-arabinose differentially activate peripheral Gr43a neurons 36
Activation of Gr43a neurons in the absence of sugar can form associative memory 40
Optogenetic activation 44

DISCUSSION 48
Attributes of sugars important for long-term memory 49
Role of gustatory receptors in long-term appetitive memory 50
Reward perception 51

PART II. ORB2'S CONFORMATIONAL SWITCH AS A BIOCHEMICAL SUBSTRATE OF MEMORY. 54
Synapse-specific molecular changes and local protein synthesis 54
How do memories outlast molecular turnover? 55
Memory encoded in protein conformations: a prion-like mechanism of memory 56
RESULTS AND DISCUSSION 58
Substitution of Orb2's prion-like domain at the endogenous locus 58
Evaluating the functional and biophysical properties of the chimeric Orb2 proteins 59
Biophysical analysis 60
Behavioral analysis 61

APPENDIX A. MURASHKA 65

Materials and Methods 76
Reward perception and long-term memory 76
Orb2's conformational switch 81
Murashka 83

Genome editing in Drosophila 85

List of Figures

Figure 1: Classification of memories by content 2

Figure 2: Flies associate the value of a sugar with odor during the two-minute training 16

Figure 3: Flies' immediate preference for a sugar is not predictive of their long-term memory 19

Figure 4: L-arabinose, a natural non-nutritious sugar, and its isomer D-arabinose separates immediate appeal and long-term reward value of a sugar. 20

Figure 5: Specificity of L-arabinose memory 22

Figure 6: Though both are sweet, flies prefer D- over L-arabinose. 25

Figure 7: D-arabinose detection does not rely on any single receptor or set of neurons. 27

Figure 8: Gr43a and Gr61a neurons and receptors are involved in L-arabinose detection. 29

Figure 9: Gr61a-expressing neurons are involved in D- vs L- preference. 30

Figure 10: L-arabinose and D-arabinose memory 32

Figure 11: Expression patterns of Gr61aGAL4 and Gr43aGAL4 35

Figure 12: Peripheral Gr43a neurons are critical for L-arabinose memory and activated differentially by D- and L-arabinose. 37

Figure 13: Further activity of Gr43a neurons in response to D- and L-arabinose 38

Figure 14: Activation of Gr43a neurons is sufficient for appetitive-associative memory. 41

Figure 15: dTrpA1 activation of Gr43a neurons does not produce long-term memory 43

Figure 16: Optogenetic activation of GrGAL4 neurons to create illusory appetitive memories 45

Figure 17: Orb2's genomic locus 57

Figure 18: The spectrum of protein aggregates 58

Figure 19: Details of the RNA-binding assay 60

Figure 20: FRAP recovery of the chimeric Orb2 proteins 61

Figure 21: FUS's disordered domain and Sup35's prion-like domain can substitute for Orb2's prion-like domain. 62

Figure 22: Murashka shows high amyloid-type vibrations in FTIR 65

Figure 23: Candidate prion-like proteins exhibit low FRAP recovery 65

Figure 24: Murashka contains a predicted disordered domain 66

Figure 25: RING domain structure 68

Figure 26: Murashka mutants have impaired long-term memory 70

Figure 27: muraGAL4 expression pattern 71

Figure 28: Generation of muraGFP and muraHA flies 73

GENERAL INTRODUCTION

Memory is a faculty fundamental to human behavior, from routine daily events to the most remarkable human achievements. Eradicating smallpox, landing humans on the moon, Beethoven's symphonies, and Borlaug's wheat all relied on the ability to accumulate and store information which, at some later time, would be retrieved to influence thoughts and behavior.

Definitions of memory understandably vary by setting. The subjective experience fits William James' description: "An object which is recollected, in the proper sense of that term, is one which has been absent from consciousness altogether, and now revives anew" (James, 1890). Our experience of memory involves ideas and events that weren't in our thoughts a mere second before we remembered them. The neurology clinic, meanwhile, deals with defects in memory, and, by carefully delineating various types of memory loss, has for decades advanced our understanding of human memory (Kofinas, 2015; Squire and Wixted, 2011). The clinical aim is to identify memory loss—amnesias—and differentiate those that impair new or recent memories ("Do you know where you are?" "Do you know what day it is?") from more extensive deficits ("Who was President when you were born?" "What year did 9/11 happen?"). The neurological exam's working definition of short-term memory is therefore the ability to hold in one's head, for a few minutes, a short series of names or objects. Long-term memory is more robust, and necessary for situating oneself in the world. It is correspondingly desolating when lost.

Because different types of memory seem to use different circuitries—impairment of one type does not universally co-occur with impairment of the other—memories are subdivided into declarative memory (facts and events you could relate to someone) and nondeclarative memory (skills and habits you pick up over time but may have trouble verbalizing, as well as simple classical conditioning that has shaped how you to respond to the world) (Figure 1) (Goldman, 2000). Nonetheless, in all these cases, memory consists of identifying features found in the world, retaining that information even when it isn't presently held in mind, and recalling that information, consciously or nonconsciously, in a way that guides later thoughts and behavior.

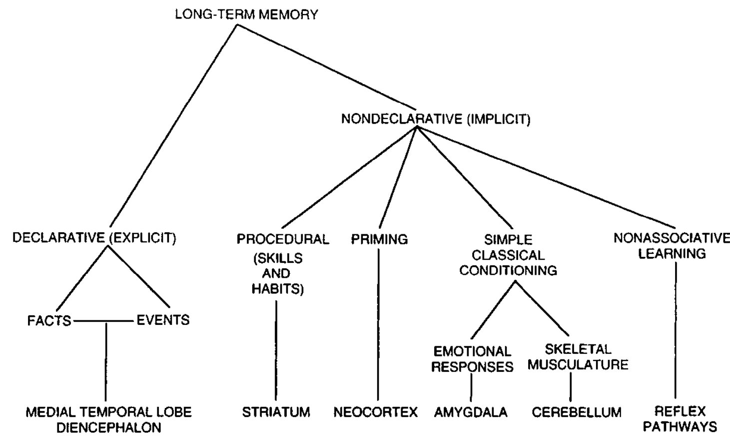


Figure 1: Classification of memories by content
Human memory is subdivided into declarative and nondeclarative memory; these different types of memory rely, at least for a certain amount of time, on the different brain structures listed at the bottom. Declarative memory is often what is referred to by common uses of memory—consciously explicit memories like your wedding day, that The Warmth Of Other Suns is the best book you've read in the past five years, that the baby's room is decorated in grey, yellow, and white. Nondeclarative memory consists of behaviors that you have picked up by repetition and experience, associations that you've nonconsciously identified, and various kinds of nonassociative learning including habituation. Taken from (Milner et al., 1998).

Neurobiological definitions of memory are slightly different: they rest on recent advances in our understanding of the cellular and molecular events involved in memory formation. Despite pronouncements on memory stretching back millennia (Aristotle, 350 BC), the neurobiology underlying memory has only in the last century begun to be addressed. The changes that take place in our brains upon memory formation are now widely considered to be changes in the ease of communication at specific synapses, between specific neurons, in specific brain regions (Kandel, 2001). The difference between remembering the Chicago Symphony playing Mahler 1 and remembering the face your daughter was making that day at lunch lies in which neurons, which pathways, these experiences modify—but the modifications that occur are presumed to be similar. These types of modifications were first postulated more than a century ago, when Ramon y Cajal described dendritic spines—these small protrusions on dendrites are what we now know to be the sites of synapses. Despite capturing images of single neurons in isolation, Ramon y Cajal suggested that these were the sites of contact between neurons. He went on to propose that the growth of new connections or the retraction of existing ones—as well as the growth or retraction of whole dendrites—could underlie activity-dependent changes in cognition (DeFelipe, 2006; Ramon y Cajal, 1892).

From a wider biological perspective, memory is important for the survival and reproductive success of a wide variety of animals, especially long-lived animals. The most basic aspects of memory, therefore—what to remember, when and how to remember, and the elemental cellular and molecular basis of memory formation and maintenance—can be studied in a variety of animal model systems. After a general introduction to the neurobiology of memory, Part I of this dissertation details our work on the Drosophila sensory system and those components necessary and sufficient for forming appetitive memory. We find, intriguingly, that the rewards an animal likes most in the moment are not always those that generate the most reliable memories. Part II discusses work focused on the molecular processes that must occur for a memory to be formed. The aggregation of Orb2 has been shown to be one of the molecular processes critical for the persistence of long-term memory, but the properties of that aggregation remain unclear and are the focus of this later section.

The cellular and molecular basis of memory

The earliest investigations into the cellular and molecular basis of memory—what actually happens inside neurons that could facilitate the changes underlying memory—began with a search for neuronal responses that were modified following learning. This work often used simplified preparations in which durable changes in neuronal activity could be generated by electrical stimulation (Kandel and Spencer, 1968). These changes, it was hoped, could offer insight into the similarly durable changes that must take place when any kind of learning generates a lasting memory. Soon after, by recording from populations of neurons in the rabbit hippocampus—a structure known from amnestic human patients to be required for long-term memory formation—long-term potentiation was identified (Bliss and Lomo, 1973; Scoville and Milner, 1957). Long-term potentiation involves repeatedly activating one neuron in a circuit in a manner that leads to a persistently greater response in downstream neurons, due to enhanced synaptic communication (Bliss and Lomo, 1973). But because a microscopic section of rabbit hippocampus still holds thousands of neurons, the exact structural and molecular changes that were responsible for this potentiation remained inaccessible. (Synaptic strength, in this context, refers to the injection of electric current into one neuron and measuring the depolarization that this current induces in the neuron's synaptically connected neighbor. Increased synaptic strength—as in the case of long-term potentiation—means that given the same amount of current injected into the upstream neuron, the downstream neuron experiences greater depolarization—the synapse between the two neurons now passes along a stronger signal than it did before. A sustained change in the opposite direction, reducing the

efficacy of synaptic communication, is another form of synaptic plasticity and is referred to as long-term depression [LTD]. Synaptic plasticity merely refers to this ability of synapses to change their communication strength.)

The potentiation phenomenon was also observed in isolated Aplysia (sea slug) neuronal circuits, but called long-term facilitation. In a sea slug, memory can mean modifying a simple innate behavior by repeatedly exposing the animal to innocuous or aversive stimuli. For example, the gill reflex can be conditioned so that a familiar touch no longer provokes withdrawal. Alternatively, after a period of habituation, if that same touch is paired with a mild electric shock, the gill withdrawal returns in an exaggerated manner—a process termed sensitization. Given the relative ease with which its single neurons and simple neural circuits could be isolated and maintained, Aplysia would go on to offer decades of experimental insights into the molecular and cellular mechanisms that modify synaptic strength (Castellucci and Kandel, 1976; Kandel, 2001). Given the wide variety of theories concerning memory storage—chemical gradients, electrical fields, base-pair changes in DNA or RNA, chains of self-reexciting neurons, or growth of new synaptic connections—the virtue of using Aplysia was the ability to examine the cellular events underlying these basic learning behaviors. In fact, it was in Aplysia that Cajal’s ideas about the synaptic storage of memory first found experimental support.

These simple instances of learning in Aplysia were found to require a circuit of just three neurons: (1) the sensory neuron that detects the touch, which forms synapses with (2) the motor neuron responsible for gill withdrawal, and (3) an interneuron that forms a synapse at the sensory neuron’s presynaptic terminal. For habituation, the sensory neuron was found to release less neurotransmitter, leading to reduced depolarization of the motor neuron (Castellucci et al., 1970; Kandel, 2012). In the case of sensitization, enhancement of neurotransmitter release was mediated by the interneuron, which releases serotonin onto the presynaptic terminal of the sensory neuron. Serotonin’s actions caused the sensory neuron to release more neurotransmitter onto the motor neuron, enhancing depolarization. By releasing more or less neurotransmitters into a synapse, the presynaptic neuron causes greater or lesser depolarization of the postsynaptic neuron, which directly influences the likelihood of action potentials. These synaptic changes in Aplysia can last anywhere from a few minutes to several days—the length of time depends, like so many human tasks, on how much training the sea slug receives. This neural potentiation, described in more detail below, is, in isolated neurons, sea slugs, and the human brain, responsible for converting short-lived neuronal activation into lasting synaptic modifications. The durable changes in synaptic strength seen in the sea slug’s simple neural circuit are, in a different context, the durable changes that we think underlie all memory (Bliss and Lomo, 1973; Castellucci and Kandel, 1976;

Kandel, 2001). While the exact relationship between the long-term potentiation [LTP] and what we understand as memory remains a topic of discussion (Martin et al., 2000; Stevens, 1998), one powerful argument for a connection is that of the several dozen mutants known to improve memory, nearly all of them also improve LTP (Lee and Silva, 2009). Other supporting findings show that saturating a neural circuit with LTP-generating electrical stimulation impairs spatial learning, while electrically evoked potentiation of auditory pathways can enhance the response of the pathway to natural sounds (Moser et al., 1998; Rogan and LeDoux, 1995).

In addition to classification by their contents (declarative versus nondeclarative), memories are also of course classified by their duration—short term, intermediate term, and long term. But because a variety of organisms are used to study memory (flies, mice, monkeys, and so on), it is difficult to define a certain period of time as short- or long-term memory. Remembering where you parked your car 24 hours ago is a fairly trivial memory, forgotten soon after, but for *Drosophila* a 24-hour memory lasts a substantial fraction of their life, akin to many months in humans. Key aspects of memory shared across all species, though, are the molecular requirements of memory's phases, requirements that are also reflected in isolated neuronal models of memory (Kandel, 2001). For short-term memory, on the order of minutes to hours, increased synaptic strength is accomplished by proteins already present; in response to neuronal activation, these proteins are frequently moved to other cellular locations (Lisman, 2003). For example, phosphorylation of neurotransmitter receptors in the membrane changes the membrane excitability of neurons (as does putting more receptors into the membrane), while modification of the exocytosis machinery increases the amount of neurotransmitters released (Shepherd and Huganir, 2007). For longer-lasting memory, persistent increases in synaptic strength further require new mRNA and protein synthesis, and therefore a signal must be sent back to the nucleus to trigger new transcription and translation (Kandel, 2001). These new proteins are responsible for stabilizing the modified synapse, and collectively form the biochemical basis of memory. The process that triggers new transcription and translation has been worked out over many years, and spanned work across a number of model organisms.

DROSOPHILA'S ROLE IN MEMORY RESEARCH

Aplysia's advantage was the ability to observe the cellular correlates of memory at single-neuron resolution. But screening for possible genes involved in memory—a task that required heroic optimism to begin with—needed more manageable organisms in which the genetic and molecular basis of these events could be studied. Insect models, especially *Drosophila*, gave researchers the ability to study memory using molecular genetics—mutations that affected memory could be identified and their consequences

characterized. Further, besides molecular genetics, flies demonstrated various kinds of behavioral memory that could easily be assayed, and they've recently begun to offer insights into circuit-level phenomena as well (Guven-Ozkan and Davis, 2014; Krashes et al., 2009).

Early work in insects pointed to memory as a polygenic trait, though quantitative trait mapping for memory-related loci was prohibitively complicated (Dubnau and Tully, 1998). Seymour Benzer undertook the first genetic screen for memory in *Drosophila*, whose tractable genetics and higher-order behavior allowed large-scale efforts to identify memory-deficient mutants. Benzer's optimism was rewarded, and the first single-gene mutant shown to be defective in memory was named *dunce* (Dudai et al., 1976). *dunce* was eventually determined to be defective in cyclic AMP (cAMP) breakdown, due to a mutation in the cAMP phosphodiesterase—in these flies, cAMP was continuously elevated (Byers et al., 1981; Davis and Kiger, 1981). A second memory-defective mutant, *rutabaga*, was found to have a mutation in adenylate cyclase, the enzyme that produces cAMP. Whereas adenylate cyclase is normally activated by higher calcium levels, this mutant adenylate cyclase was not. That these first identified mutations would land in the same biochemical pathway—*rutabaga* makes cAMP while *dunce* degrades it—was quite remarkable. Further work in *Drosophila* found that blocking the activity of a cAMP-responsive transcription factor CREB (*dCREB2*) blocked long-term memory, lending support to the idea that new transcription is critical for long-term memory; short-term memory, in contrast, was not affected by inhibiting CREB (Yin et al., 1994). This coincided with work in Aplysia: by reducing CREB's ability to bind its genomic CRE sequences (by outcompeting those targets with CRE-containing oligonucleotides), long-term facilitation was blocked in presynaptic sensory neurons (Dash et al., 1990; Dash and Moore, 1996). Work in mice, using targeted mutations of CREB, found that certain CREB isoforms were important in long-term memory and LTP but dispensable for short-term memory, again pointing to a role for cAMP-prompted transcription in long-term memory (Bourtchuladze et al., 1994).

Initial work in invertebrate neurons identified synaptic facilitation as occurring largely in presynaptic neurons—enhanced release of synaptic vesicles provoked greater depolarization of the postsynaptic neuron. Early work in mice, by contrast, focused more on postsynaptic mechanisms found in the hippocampus—modifying and adding postsynaptic (typically AMPA) receptors allowed a certain level of neurotransmitter release to generate greater depolarization (Glanzman, 2010). Subsequent studies, however, demonstrated postsynaptic mechanisms of long-term potentiation in a variety of invertebrates, while presynaptic mechanisms of LTP have been described in vertebrates (Glanzman, 2010).

A picture that combines these molecular events begins in the presynaptic neuron, where the binding of serotonin to its G-protein coupled receptor activates adenylate cyclase to increase production

of cyclic AMP (cAMP), and, consequently, the activity of cAMP-dependent protein kinase A (PKA). This will lead to greater neurotransmitter release onto the postsynaptic neuron, but it will also lead to the phosphorylation and activation of a set of transcription activators, the most famous of which is the above-mentioned CREB (Kandel et al., 2014). The most well-characterized target of CREB, C/EBP, is itself a transcription factor that promotes the transcription of certain target genes (Alberini, 2009; Alberini et al., 1994; Kandel, 2012). The complete list of activity-dependent CREB-responsive genes remains in progress, but includes many known to be involved in synapse growth and development, including neurotrophic factor BDNF; synaptotagmin IV, a regulator of synaptic vesicle release; and the synaptic vesicle transporter VMAT (Benito and Barco, 2010). In this way, synaptic activation, initially observable as a local and transient influx of calcium ions, eventually signals back to the nucleus and prompts the production of new mRNAs and proteins required to maintain the newly altered synapse. Postsynaptically, neurotransmitter binding depolarizes the postsynaptic neuron; simultaneous glutamate binding and depolarization opens NMDA receptors (by displacing an obstructive Mg2+ ion) that allow calcium to flow into the postsynaptic neuron. Increased calcium levels activate a variety of kinases, including Ca2+/calmodulin-dependent protein kinase II (CAMKII) and PKA; PKA both promotes modification of synaptic AMPA receptors and places additional receptors into the postsynaptic membrane, all while activating new transcription and translation (Glanzman, 2010).

All these molecular events, triggered by certain kinds of neuronal activation, combine to stabilize synaptic changes and together form the biochemical basis of memory.

Where are memories located?

But given the complexity of the human brain (and still ~100,000 neurons in the *Drosophila* brain), there exists a disconnect between the psychological phenomenon of memory and the molecules that are responsible for it. We know what recalling a memory feels like, and we have the beginnings of a sense for the molecular changes that explain persistent synaptic modification, but in between those sits 85 billion neurons. Understanding the formation, storage, and recall of a long-lasting memory ultimately requires understanding where in the neural circuitry of the human brain these molecular changes occur. Early experiments by Lashley attempted to isolate cortical structures required for memory by making precise incisions into rats' brains before or after training, but no incision was sufficient to prevent memory formation or recall (Lashley, 1950). These experiments were interpreted to mean that memories were stored diffusely throughout the cortex. This was consistent with Lashley's further finding, that memory impairments, while not correlated with any precise incision, became increasingly apparent with

the amount of cortex removed. Similarly, we know from neurosurgery patients sixty years ago that electrically stimulating certain areas of the brain (e.g., the superior temporal gyrus) causes vivid experiences of past memories, particularly aural events. One patient said she heard White Christmas playing on the radio, while another overheard his mother saying that his brother's coat was on backwards (Penfield and Perot, 1963). There's obviously no dedicated 'White Christmas' circuit in the human brain—crude activation of cortical neurons in that region was enough to resurrect a memory, stored either in that region or in some downstream pathway. But while memories seem to be stored diffusely in the cortex, there is in fact a small brain structure required for the formation of spatial and declarative memory: the hippocampus. The first indication of this famously came from a human patient, H.M., whose epilepsy was cured but ability to form new memories lost by surgery that removed (most of) both hippocampi and nearby structures (Scoville and Milner, 1957). Though we now know that memories are indeed formed in a manner that requires the hippocampus, after an extended period of time memories do begin relying less on the hippocampus and more on the cortex—a process termed consolidation (Preston and Eichenbaum, 2013). But identifying exactly which neurons in the brain are using the canonical CREB and PKA machinery for memory A, and whether some, all, or none of those same neurons (and perhaps synapses) are used for related memory B—and what process decided that A would use A's neurons and not B's?—remains a formidable challenge.

If our current thinking about synaptic plasticity is correct, it will be very difficult to prospectively isolate, observe, or manipulate the select neurons and synapses involved in a memory. Recent experiments have biased memories to use certain neurons (by mosaically overexpressing CREB), and others have even labeled hippocampal neurons activated during a memory's formation (Han et al., 2007; Liu et al., 2012b; Tonegawa et al., 2015). These indicate that it is possible to selectively manipulate enough neurons to repress or drive the expression of certain memories, but both still label a relatively large number of neurons. Identifying which neurons in the brain are responsible for which memories, and why, will ultimately involve knowing which out of 85 billion neurons have undergone certain molecular changes, and why the molecular changes happened in those neurons but not their neighbors.

But before these molecular changes occur, there must be some neuronal event that prompts them. If the memory is caused by an external stimulus, the signal will originate somewhere in the sensory system. If one is able to trace the path of neural activation, from the origins of the sensory signal to the neurons whose synapses undergo modification, all the way to motor neurons that execute the appropriate behavior, one can comprehensively describe that memory. The role of the sensory system in determining

what information should be passed along, and how that information leads to memory formation, is the topic discussed next.

SPECIFIC OBJECTIVES OF THE DISSERTATION

The first and largest part of this dissertation contains work focused on understanding the initial steps of memory formation: activity in the sensory system. What is it about a reward that triggers memory formation? And which sensory neurons are critically involved in the perception of a reward? The answers help explain how a fleeting experience can be quickly sensed by the animal, and how that perceptual process helps dictate the likelihood of memory formation. The eventual result of memory formation, however, will—as described above—be an alteration at a limited number of synapses several neurons later. The second part of the dissertation focuses on the molecular mechanisms that transform this sensory information into a stable memory.

PART I. REWARD PERCEPTION AND LONG-TERM MEMORY

A central question in long-term memory work is why animals remember some but not all experiences. What features of an experience are critical for generating long-lasting memory, and how does the animal recognize these essential features? The relationship between experiences and memory has traditionally been investigated in humans and primates due to our ability to report what we have and haven't perceived. As mentioned above, however, there are certain elemental aspects of higher-order processes that can meaningfully be addressed in animal models, including *Drosophila*.

While definitions of memory vary depending on the level of description, in all cases the first requirement for remembering an external event is a sensory response. If an event is to be remembered, it must first be detected, even if subliminally. If your sensory neurons don't have receptors that can recognize a certain kind of impulse, you will never be aware that one has occurred, let alone remember it. In the case of humans, this is true for very large and very small physical events, most sound frequencies and wavelengths of light, and of course much else. Relevance, in this case, is largely genetically programmed into organisms, with evolution having dictated the kinds of stimuli we can perceive. Quite apart from the central neurons involved in memory then, understanding which phenomena animals remember—and why they're remembered—is a question first to be asked at the level of sensory neurons. It is sensory neurons that are the interface between an animal and its surroundings, and they help dictate whether an event will be perceived and remembered.

Even within the set of easily perceivable experiences, though, only some are remembered; most are not. What aspects of an experience help determine whether it will be remembered? Of course, many factors contribute: animals often must be motivated to remember (e.g., if the reward is food, they typically need to be hungry) and the experience must be relevant to the animal at that moment. But very similar sensory stimuli, when all else is held equal, can vary in the strength of memory they generate. It is these cases that help shed light on the role that sensory neurons play in relaying information that may or may not be remembered. Studying sensory neuron involvement in memory selectivity is currently more feasible in organisms and sensory systems with a limited number of sensory neurons. In gustatory systems, sensory neurons behave more or less in labeled line fashion, with activity of sensory neurons faithfully transmitted to higher brain areas (Barretto et al., 2015; Harris et al., 2015; Liman et al., 2014). The properties of the fly gustatory system can therefore offer an opportunity to study why similar rewards—in this case sugars—lead to different levels of long-term memory, and the role that sensory neurons play in dictating that memory formation.

Reward components

Because it is the perception of an experience that will help determine whether we remember it, reward perception is fundamental to all aspects of animal behavior influenced by reward learning. Rewards are stimuli that are intrinsically attractive to the organism, prompting approach behaviors (though what rewards consist of obviously varies depending on the organism studied) (Hu, 2016). However, even though rewards can be a single object, the psychological processes underlying the experience of a reward are several-fold. First, a reward is, in the moment, 'liked'—the experience of the reward is appealing to the organism while the experience is happening (Anselme and Robinson, 2016). This can be evaluated by a number of different measures—which food an animal chooses to eat, the stereotyped facial responses of mice or humans when eating something appetitive (Anselme and Robinson, 2016). 'Wanting', in contrast, is the reward's ability to motivate the animal to pursue the experience again, even when the reward is not currently present. 'Wanting' can be measured by how hard subjects are willing to work, or how long they are willing to persevere, to obtain a reward. While liking and wanting generally co-occur, one can imagine that there may be rewards liked in the moment that nevertheless do not generate strong motivational urges, or vice versa. In fact, these two processes appear to be separated in the brain—at least the rodent brain—with 'liking' carried out by opioid and endocannabinoid systems, while 'wanting' is mediated by dopaminergic systems (Anselme and Robinson, 2016). The relationship between these two aspects and reward learning, if any, however, remains unclear (Berridge and Kringelbach, 2015; Berridge et al., 2009).

The basic model for reward learning is termed associative conditioning: the subject learns that an arbitrary event tends to precede either a positive or negative experience. Also known as classical conditioning, it causes the subject to associate the predictive or conditioned cue with the unconditioned stimulus (recognizing the value of the unconditioned stimulus requires no training—think of sugar when an animal is hungry, or, most famously, the tale of Pavlov's dogs). Operant conditioning, in contrast, relies on the subject performing some action that then has a learnable effect—a mouse receiving food when it presses a bar, an enclosed chamber that heats up when the fly crosses to the wrong side.

Associative conditioning is one of the ways to test the role of perception in memory. Both the unconditioned stimulus and the conditioned stimulus can be systematically modified to determine features that the sensory and memory circuitry deem important—eliminate those aspects, and the animal will no longer remember the association. If, say, food is being used as a reward, and food can impart both rewarding tastes and rewarding nutritional value, is either one of those sufficient to generate memory? Does memory formation require both? This can be addressed using rewards that contain one aspect of the

dual-natured reward but not both. If one is using model organisms whose sensory circuitry can be altered, this enables an assessment of how perceptions are generated. If, when a subset of neurons is silenced, the animal no longer recognizes the reward, or recognizes it in a weakened or modified way, the reward may no longer mean enough to the animal to generate associative memories. Those neurons would therefore be recognized as playing an important role in the perception of that reward.

It should be noted that not every associative memory paradigm works in all animals. The conditioned stimulus needs to be perceivable to the animal, of course, and the memories formed need to be testable in the lab. This associative pattern of learning is fundamental to a wide range of animal behaviors, and with increasing levels of sophistication can explain a great deal of human learning as well. But as described above, to examine the genetic and circuit-level basis for these kinds of memory tasks, as well as the higher-order processes involved in memory formation and recall, *Drosophila* is an approachable system. Single genes can be modified, single neurons or entire circuits can be silenced or eliminated, and physiological states can be manipulated. If one aims to address the sensory basis of reward memory, *Drosophila* provides key advantages, not least a tractable nervous system.

Food as a rewarding memory

In natural environments, the positive experiences an animal remembers are widely assumed to be salient to the animal and, at least in some sense, rewarding. Long-term associative memories in particular are supposed to reflect the intensity of past responses to rewards. The experiences we remember, however, are not always those we expect to remember. The connection between the immediate perception of a reward and that reward's ability to generate memories therefore remains of general interest.

Experimental investigations into memory formation necessitate the ability to create memories in the organism of study. For paradigms using positive rewards to generate associative memories, food, and in particular sweet food, has been frequently used as the reward since it provides both immediate appeal and delayed nutritional value. But that makes food a complex reward, and different animals have further evolved distinct food preferences and nutritional sources (Dethier, 1976). Just within *Drosophila*, certain species prefer rotting fruits while others mainly feed on mushrooms, cacti, or hibiscus flowers (Markow and O'Grady, 2005). Since food is always a mixture of various compounds, and feeding occurs in an environment that often includes predators, quickly evaluating potential food requires simultaneous processing of multiple complicated stimuli (Charlu et al., 2013; Yarmolinsky et al., 2009). Remembering a food source in nature thus requires quickly identifying, from a wide range of sensory information, key

cues that signal the presence of food. But how these cues are recognized, and how they quickly generate appropriate memories to guide future food-seeking behavior, remains unclear.

CHEMOSENSORY PERCEPTION IN DROSOPHILA

Drosophila's chemosensory apparatus consists of the gustatory and olfactory systems. While both are responsible for sampling a wide range of chemical compounds, these systems are arranged in quite different ways. In the olfactory system, roughly 2,500 olfactory receptor neurons each express Or83b (a required coreceptor) and one other of about 60 olfactory receptors genes (Vosshall and Stocker, 2007). This relatively large number of neurons, each of which expresses only a single olfactory receptor family member, parallels the arrangement found in vertebrate olfactory systems, though with fewer neurons and fewer receptors. In this way, incredible numbers and combinations of odorants can be identified by the combined activities of many olfactory neurons. The gustatory or taste system, in contrast, uses combinatorial expression of gustatory receptors in a more limited number of neurons, allowing each neuron to respond to a range of compounds. While bitter neurons can simultaneously express more than two dozen bitter receptors, only nine sugar-sensing receptors have been identified in *Drosophila*: Gr5a, Gr43a, Gr61a, Gr64a, Gr64b, Gr64c, Gr64d, Gr64e, and Gr64f. Sugar neurons—located on the surface of the mouth (labellum), lining the esophagus and intestine, in the brain, and on the wingtips and legs—express different combinations of these nine receptors. The sugar-sensing receptors appear to respond differently to different sugars, with Gr5a responding to glucose and trehalose (Dahanukar et al., 2001; Dahanukar et al., 2007), Gr43a responding to fructose (Miyamoto et al., 2012; Sato et al., 2011), Gr61a responding to glucose (Miyamoto et al., 2013), Gr64a responding to various polymers and sucrose (Dahanukar et al., 2007; Jiao et al., 2007), Gr64e responding to glycerol (Wisotsky et al., 2011), and Gr64f found to be a required coreceptor for responses to all sugars except fructose (Jiao et al., 2008).

Gustatory neurons run from the peripheral taste organs up to the subesophageal ganglion in the fly brain. It is clear that this area of the brain is connected to higher order structures involved in memory, but second-order neurons carrying the sensory signal to other areas of the brain remain, with a small number of exceptions, unidentified (Kain and Dahanukar, 2015; Liu et al., 2012a; Miyazaki et al., 2015). It seems likely that second-order neurons, once comprehensively identified, will provide insight into the circuit-level consequences of the differential activation of various primary gustatory neurons.

Objective of the proposed study

To examine how animals evaluate salient features of food, and how the immediate perception of a reward influences long-term memory formation, we have used a *Drosophila melanogaster* training procedure known as the associative-appetitive memory paradigm (henceforth referred to as the 'memory paradigm') that approximates food-seeking behavior (Colomb et al., 2009; Krashes and Waddell, 2011; Tempel et al., 1983). This memory paradigm is ideal for three reasons: one, it is an ethologically relevant behavior; two, the internal state of the fly (hunger) influences the likelihood of memory formation; and three, features of the sugar reward dictate both the likelihood and duration of memory.

Sugar rewards therefore provide an excellent opportunity to study the relationship between the immediate perception of an experience and that experience's ability to generate long-lasting memory. Flies find many sugars sweet (Gordesky-Gold et al., 2008), and the relationship between the propensity to eat a sugar (or, colloquially, how much the flies 'like' a sugar) and how well the sugars generate long-term associative memory can readily be addressed. Some sugars reliably produce memory, while others do so much less reliably, and still others produce no memory at all. Long-term associative memories are therefore thought to reflect the value of the rewards at the time the rewards were presented—how much the flies like the sugar, and how important the sugar is perceived to be by the flies' memory circuitry.

RESULTS

Flies rapidly evaluate the value of a sugar

In the olfactory memory paradigm, hungry flies spend two minutes in an odor-filled tube, with access to water, and then are moved to a second tube for two minutes, filled with a different odor and with access to sugar. After the four minutes of training, with only two minutes spent in the presence of sugar and the second odor, flies will remember the sugar-paired odor for days (and will run towards the odor if given the opportunity) (Tempel et al., 1983). The likelihood of memory formation, and thus the memorability of the reward, is measured by the proportion of flies that choose the previously reward-paired odor. We first wanted to determine when the flies evaluate the attributes of sugar important for memory: during the two-minute training, or over some broader time window post-training? A pair of experiments by Dus and colleagues indicated that, in the absence of taste, the time required for flies to evaluate sugars using metabolic information alone was at least fifteen minutes. Using taste-blind flies (both combined Gr5a;Gr64a-null mutants and *poxn* mutants), they found that *Drosophila* will eventually prefer eating sugars with energy value over plain agarose, but that this preference takes 15-45 minutes to develop—much longer than the two-minute training (Dus et al., 2011). This suggests that if energy-sensing is the sole cue flies are able to exploit, it is unlikely that a behavioral manifestation of this process will be seen in response to the short two-minute training. If, however, memory processes' evaluation of the sugar continues beyond the two-minute training in some way, it could imply that post-training nutritional status is what actually matters. We tested these possibilities in two ways. First, flies trained with sucrose, which is normally sufficient to produce robust 24-hour memory, were immediately placed either into vials with standard food, or into vials with only water. We reasoned that if sucrose's energy value is assessed at some point after the two-minute training, then providing rich nutritious food immediately after training may interfere with the fly's ability to attribute its now-increased energy status to the two-minute training, and thus would interfere with memory formation. However, memory scores did not differ between the flies given food and those given only water (Figure 2A). Second, we trained flies with L-sorbose (a sugar that lacks energy value and produces only minimal long-term memory) and then immediately fed them sucrose. We reasoned that if the energy evaluation occurred outside the two-minute training, feeding on sucrose may substitute as an energy cue and lead the flies to remember the odor that was paired with L-sorbose. However, there was no enhancement of memory in the flies fed immediately with sucrose (Figure 2B). These results suggested to us that the evaluation of a sugar's

energy value, thought to be critical for long-term memory of the paired odor, occurs within the two-minute training session, consistent with reports by other groups (Burke and Waddell, 2011).

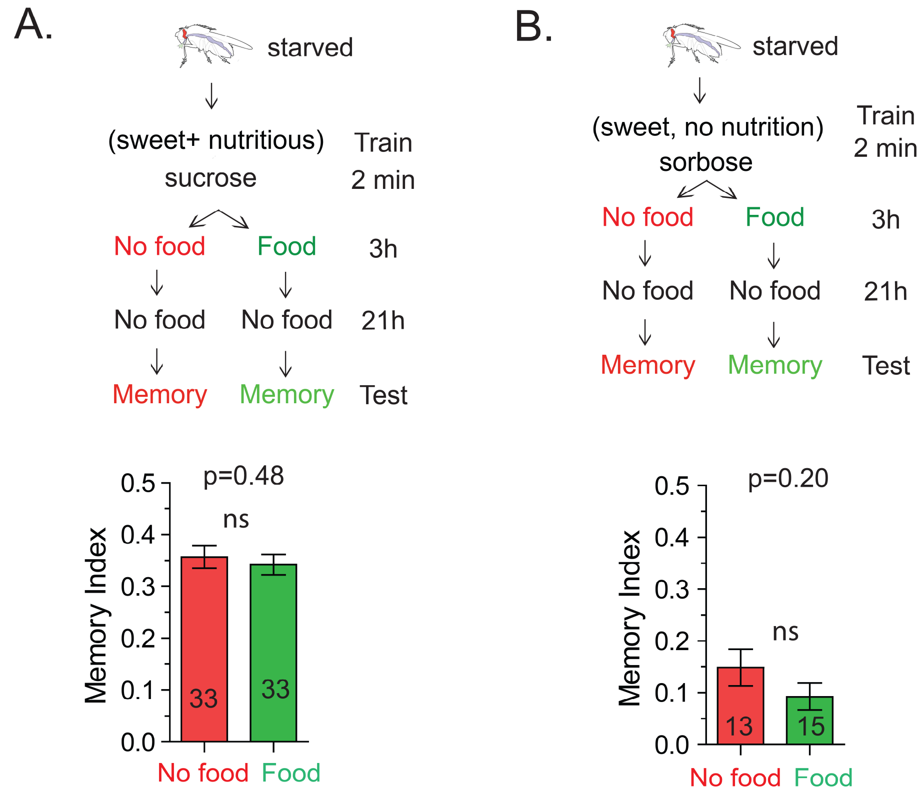


Figure 2: Flies associate the value of a sugar with odor during the two-minute training.
(A) Flies trained with sucrose, which produces robust 24-hour memory, were immediately provided either rich nutritious food or water. If sucrose's nutritional value is assessed beyond the two-minute training, nutritious food immediately after training may interfere with the fly's ability to attribute its nutritional status to the two-minute training, and thus interfere with memory formation. However, flies trained with 1M sucrose and given food or no food for 3 hours post-training show similar memory, unpaired t-test, p = 0.480. **(B)** Flies trained with L-sorbose—a sweet but non-nutritious sugar that produces short- but not long-term memory—were immediately fed sucrose. If the nutritional evaluation occurred after the two-minute training, immediate feeding on sucrose may substitute as a nutritional cue, resulting in enhanced long-term memory. However, flies trained with 1M L-sorbose and given sucrose immediately after training show similar memory to flies not given sucrose, suggesting that the critical association period was confined to the two-minute training, unpaired t-test, p = 0.207.

FLIES PREFER D-ARABINOSE BUT FORM MORE RELIABLE MEMORIES OF L-arabinose

Given that the flies' evaluation of sugars occurs within the two minutes of training, we can study the differences in flies' responses to sugars over this short period of time and compare it to the sugars' ability to generate long-lasting memory. One intuitive explanation connecting immediate perceptions and memory is that whatever components of a sugar that make it more appealing are the same components that animals find rewarding. This would predict that the more appealing a sugar is, the better the flies will remember it. Another possibility is that certain components of a sugar can reinforce memory relatively independently of the sugar's immediate appeal: they could somehow signal value that isn't necessarily recognized by the circuitries that drive approach and consumption behavior. In a complex environment, where an animal is perceiving many signals simultaneously, such processing could ensure that regardless of the immediate perception, stimuli of long-term relevance will be remembered.

Previous work has most often used sucrose as the reward, though a couple studies have used additional sugars (Burke and Waddell, 2011; Fujita and Tanimura, 2011). In order to have a wider and more ethologically relevant set of sugars to evaluate flies' preferences and memory, we trained flies with various sugars, including those that are present in *Drosophila melanogaster's* natural diet of ripening fruits (Figure 3A). L-sorbose, L-arabinose, L-rhamnose, and L-fucose are all non-nutritious sugars (Figure 3B); this allows us to avoid, in both preference and memory assays, the confounding effects of nutrition. These four sugars are perceived differently by the flies—this is evident in Figure 3C, where L-fucose is the most-preferred sugar, followed by L-arabinose, then L-sorbose. We also used these sugars as rewards in the memory paradigm, along with D-arabinose, an isomer of L-arabinose. We observed that the relative appeal of a sugar does not always predict its ability to act as a rewarding stimulus for long-term associative memory (Figure 3B-E). That is, just because a sugar is more preferred does not mean it will be better remembered This was apparent for comparisons of multiple sugars—e.g., L-fucose is greatly preferred over L-arabinose, while if anything L-arabinose generates slightly better memory—but nowhere so striking as the difference between two isomers, D- and L-arabinose (Figure 4A). Neither D- nor L-arabinose is nutritious (Figure 4B) though both taste sweet to flies (Figure 4C): when either is paired with plain agarose, flies eat the sugars. Flies overwhelmingly preferred D-arabinose to L- (Figure 4D), and form similar short-term memories of both (Figure 3D). However, it is L-arabinose, not D-, that is more effective in producing long-term memory (Figure 3E and Figure 4E). That is, when L-arabinose is paired with an odor, 24 hours later flies will prefer the L-arabinose-paired odor more than they would if that same odor had instead been paired with D-arabinose. This relative ineffectiveness of D-arabinose in producing long-term memory is consistent with other studies—other groups have used up to 3M D-arabinose to train flies, and long-term memory of it is always significantly lower than sugars like sucrose,

and in some cases not different from zero (Burke and Waddell, 2011; Cervantes-Sandoval and Davis, 2012; Fujita and Tanimura, 2011).

THE BEHAVIORAL DIFFERENCE BETWEEN D- AND L-ARABINOSE IS IN THE CHEMICAL NATURE OF THE SUGARS

We first wondered whether these behavioral differences are due to the high concentration (1M) of sugars, though 1M to 3M sugar is standard in memory assays (Burke and Waddell, 2011; Cervantes-Sandoval and Davis, 2012; Yamagata et al., 2015). The preference for D-arabinose persisted even when the sugar concentrations were reduced 100-fold (Figure 4D). 10mM D-arabinose was still preferred over 10mM L-arabinose. Preference for D-arabinose began to shift only when the D-arabinose concentration was reduced to less than a third of the L-arabinose concentration (Figure 4F). A similar difference in sensitivity to D- and L-arabinose has been reported in the blowfly *Phormia regina*, where the D-arabinose taste threshold is about 5 times lower than that of L-arabinose (Hassett et al., 1950)—that is, *Phormia* was much more sensitive to D-arabinose than L-arabinose.

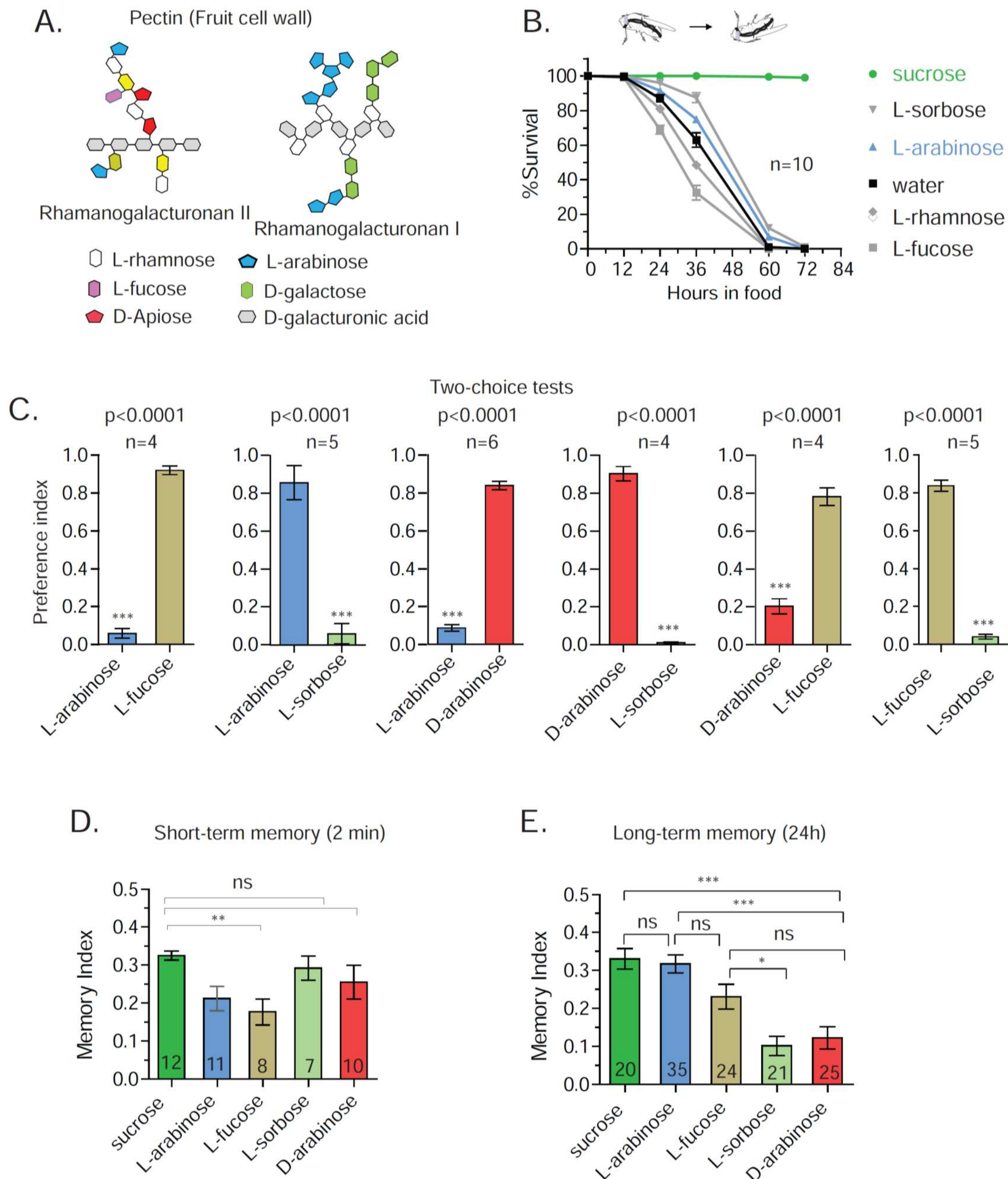


Figure 3. Flies' immediate preference for a sugar is not predictive of their long-term memory.
(A) Schematic of selected pectic polysaccharides present in fruits' cell walls, adapted from Harholt et al. (Harholt et al., 2010). **(B)** Survival curve of flies given only 1M sucrose, L-sorbose, L-arabinose, L-rhamnose, L-fucose, or water, n = 10 for each time point (50 flies per n). **(C)** Flies' immediate relative preferences as measured by two-choice assays. Flies preferred L-fucose > D-arabinose > L-arabinose > L-sorbose. **(D)** Short-term (2 minute) associative memory scores for the sugars. For short-term memory, D-arabinose = L-sorbose ≥ L-arabinose = L-fucose. **(E)** Long-term (24 hour) associative memory scores for the sugars. The order of long-term memory score: L-arabinose ≥ L-fucose ≥ D-arabinose = L-sorbose. L-fucose is a component of pectin as well, though the amount is low compared to L-arabinose. Results with error bars are means ± s.e.m. ns, not significant. *≤0.01, **≤0.001 and *** ≤0.0001.

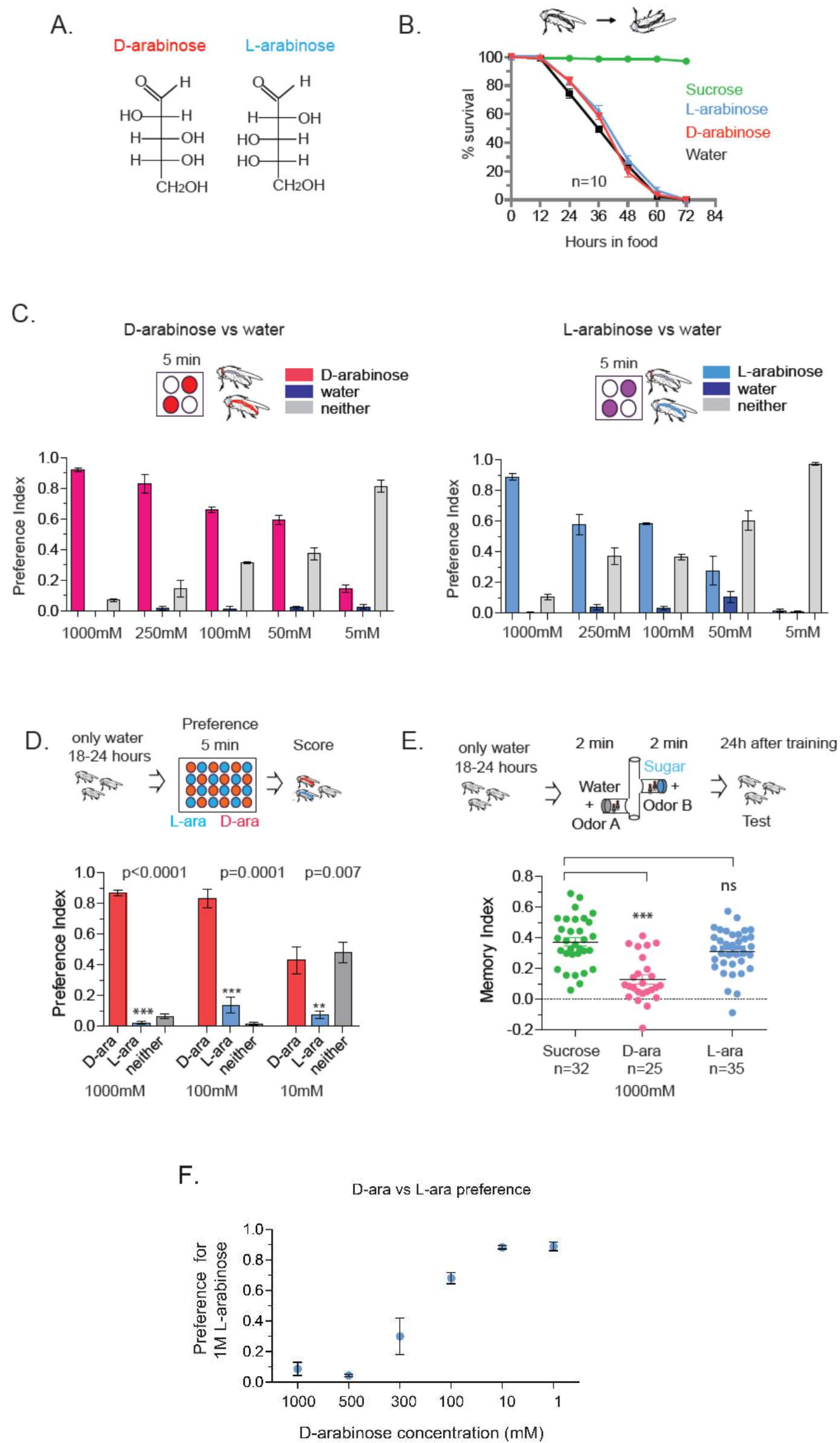


Figure 4. L-arabinose, a natural non-nutritious sugar, and its isomer D-arabinose separates immediate appeal and long-term reward value of a sugar.

(A) Structures of D- and L-arabinose. **(B)** Survival percentages for flies given solely 1M sugar solutions. n = 10 (50 flies per n) for each time point. **(C)** Two-choice feeding assays pairing D-arabinose with water and L-arabinose with water; likelihood of detection and consumption fall with decreasing concentration. **(D)** Two-choice tests comparing flies' preference for D- and L-arabinose when both sugars are presented side-by-side for 5 mins. n = 4 (50 flies per n). **(E)** Long-term (24 hours) memory of sucrose and D- and L-arabinose. **(F)** At equal concentrations, flies overwhelmingly prefer D-arabinose to L-arabinose; the preference shifts when L-arabinose is between ≥ 3 times more concentrated than D-arabinose. Results with error bars are means ± s.e.m. ns, not significant. *≤0.01, **≤0.001 and *** ≤0.0001. The data in Figure 1C was analyzed by unpaired two-tailed t-test and Figure 1D and 1E were analyzed by one way ANOVA with Bonferroni's multiple corrections. The numbers within the bar graphs indicate n; detailed explanations of what constitutes a single n is found in Materials and Methods.

We next looked to determine whether the memory seen with L-arabinose could be explained some other way, since previous groups had suggested that non-nutritious sugars were unable to generate strong long-term memory. (Burke and Waddell, 2011). The ability of L-arabinose to produce long-term memory was not restricted to sugars obtained from one particular source (Figure 5A)—suggesting that contamination of L-arabinose with some other nutritious sugar wasn't the cause—nor was it restricted to the particular wild-type flies used in the experiment (Figure 5B), or the particular experimenter, meaning that the strong memory seen with L-arabinose was reproducible. Because we saw L-arabinose memory in all these conditions, we were more confident that the memory was real and robust. A polymer of L-arabinose and galactose, arabinogalactan, did not generate long-term memory, nor did the natural L-sugar L-rhamnose (Figure 5C). It was conceivable that flies would have generically better memory at a certain time of day, perhaps artificially boosting L-arabinose memory scores, but the time of day that training occurred did not affect memory of L-arabinose (Figure 5D). Bacteria are known to use L-arabinose (Watanabe et al., 2006), and so we wondered whether the bacteria modified the sugar somehow and then released it in a form that the fly could perceive as nutritious—thereby causing the flies to remember a 'nutritious' L-arabinose—but a cocktail of broad-spectrum antibiotics for the two days prior to training had no effect on L-arabinose memory (Figure 5E).

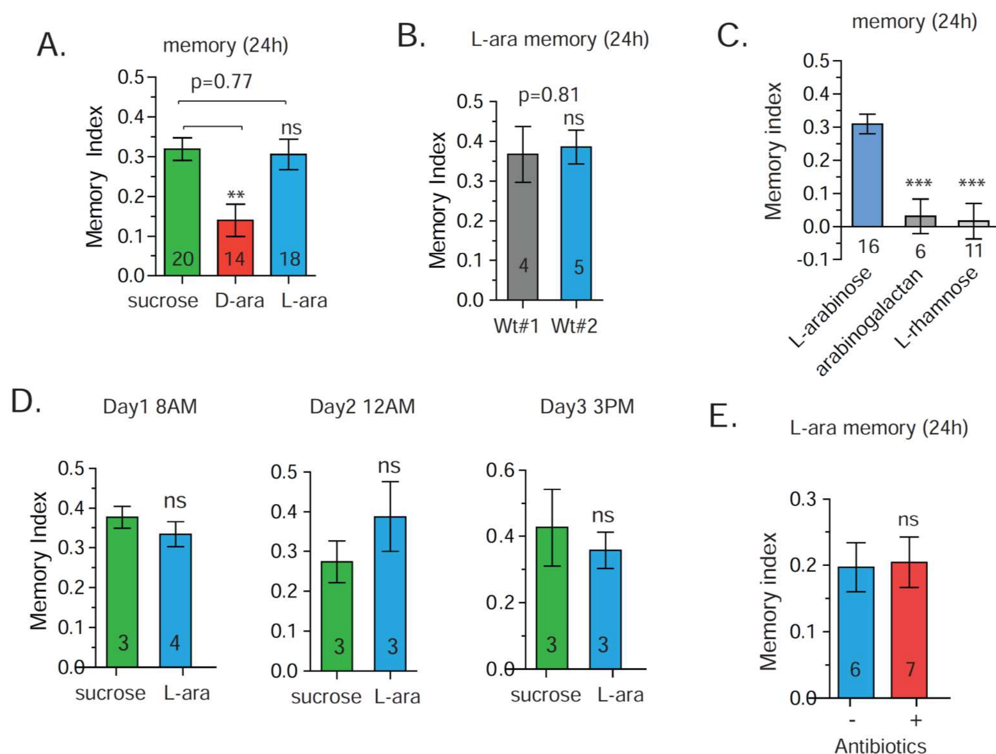


Figure 5. Specificity of L-arabinose memory.
(A) L- and D-arabinose from an independent source (USB) generated long-term memory equivalent to those sourced from Sigma (Figure 3E), controlling for other contaminating sugars. **(B)** Canton-S flies obtained from two different labs show similar long-term memory when trained with L-arabinose. **(C)** While L-arabinose forms memory, an L-arabinose-galactose polymer does not; neither does another natural L sugar, L-rhamnose, demonstrating the selectivity of L-arabinose memory. **(D)** Since flies' feeding behavior can be influenced by time of day, flies were tested at different times on consecutive days. Under all testing conditions the flies were able to form long-term memories of L-arabinose. **(E)** Flies fed a cocktail of three broad-spectrum antibiotics (kanamycin, ampicillin, and tetracycline) for the 48 hours before training show memory indistinguishable from untreated control flies. Results with error bars are means ± s.e.m. ns, not significant, *≤0.01, **≤0.001 and *** ≤0.0001.

When we measured consumption over a range of concentrations both by Capillary Feeder (CAFÉ) assays (Figure 6A) and by mixing radioactive ^{32}P in the sugar during two-choice assays (Figure 6B), flies consumed more D- than L-arabinose. The above-mentioned two-choice test (using microtiter plates) gives a snapshot of how the flies ate over the previous five minutes, which is a good general indication, but the capillary feeding assay provides a much finer measurement of the volume the flies consume. This was an assay that could tell us, in real time, how much D-arabinose or L-arabinose the flies were eating. The results from the capillary feeding assay were again consistent with flies liking

D-arabinose more (Figure 6A); flies ate more when given D-arabinose than they did when given L-arabinose. Over time, however, flies consumed less D- and L-arabinose than nutritious sugars (data not shown), consistent with other studies (Dus et al., 2011; Stafford et al., 2012): consumption reached a plateau after ~14 minutes for D and ~30 minutes for L (Figure 6A). Post-ingestive cues have been found to influence behavior approximately 30-45 minutes after feeding, which is quite possibly a reason flies stopped consuming the nutritionless D- and L-arabinose (Dus et al., 2011). However, consumption alone is not necessarily an indicator of whether animals like one sugar more than the other. If the difference in consumption is indeed due to their preferring one sugar more, then the animal should spend more time in the preferred sugar. We therefore monitored by video the behavior of single flies as they fed on colorless side-by-side D- and L-arabinose solutions (Figure 6C) and observed that over time they spent much more time on D-arabinose than L-, consistent with the higher overall consumption. These time differences are not due to differences in mere detection of the 1M D- and L- arabinose: detection rates were very similar at high concentrations and began to differ only when concentrations were dropped to ≤ 50mM (Figure 4C).

Curiously, the proboscis extension response (PER)—when the fly's proboscis is touched with a sugar, it instinctively extends, thereby reporting immediate acceptance of a taste stimuli—was similar between D- and L- arabinose over a range of concentrations (Figure 6D), consistent with other reports that PER depends more on the intensity than chemical nature of the sugar (Masek and Scott, 2010; Stafford et al., 2012). This PER behavior essentially asks whether the fly finds the substance attractive, which both D- and L-arabinose are. But we know that proboscis response does not involve ingestion of the sugar, and even detecting a sugar well enough to ingest it is not always sufficient for long-term memory. For example, a choice test between water and concentrations of sucrose showed that there was no difference in the likelihood of consumption from 1M all the way down to 10mM sucrose; only when sucrose concentration was dropped to 1mM did flies have trouble detecting and eating sucrose (Figure 6E). However, sucrose concentrations had to be ≥ 100mM to reliably produce long-term memory (Figure 6F).

Although in memory assays the flies are exposed to sugars for just two minutes—a time frame where they seem not to consume a great volume of sugar—it remains possible that consuming too much non-nutritious sugar, such as D-arabinose, is somehow a negative reinforcement. We think this is unlikely for two reasons: first, L-fucose, also non-nutritious, was preferred by flies even more than D-arabinose, and produced better long-term memory (Figure 3C, 3E). Second, at lower concentrations of D-arabinose, where the flies could still detect D-arabinose but consumed less of it, there was no increase in memory

(Figure 6G); like other sugars, lowering D-arabinose concentration also reduced the memory score, indicating that it is not the case that D-arabinose would be remembered better if only the flies ate less sugar. There is therefore no obvious relationship between how much of a non-nutritious sugar is consumed and its ability to act as a rewarding stimuli for long-term associative memory. Taken together, these results suggest that two chemically similar sugars can elicit quite distinct levels of immediate and long-term behaviors, and that immediate behavioral responses are not always predictive of long-term behavioral responses: while flies find D-arabinose more immediately appealing, L-arabinose is more salient for long-term memory.

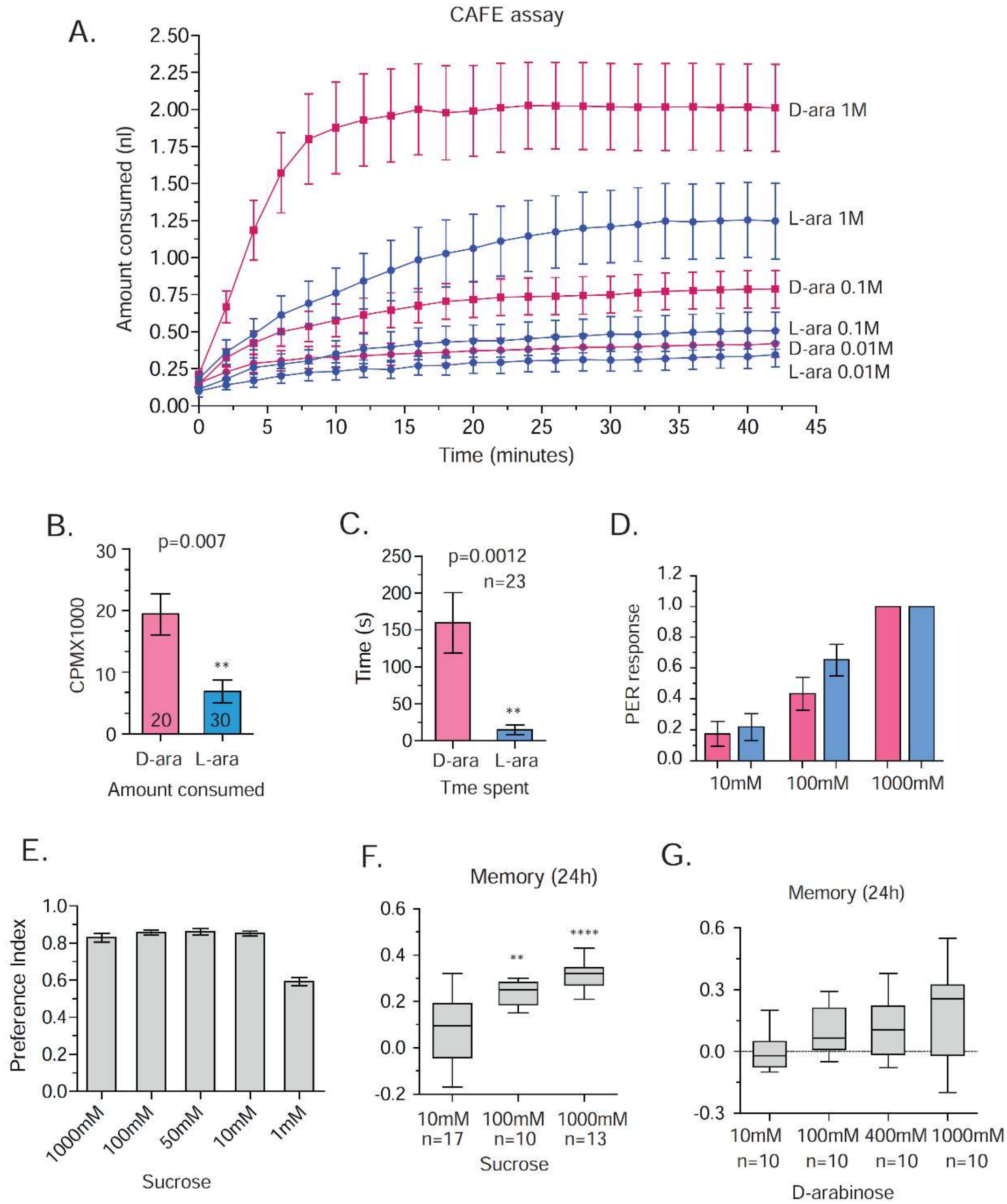


Figure 6. Though both are sweet, flies prefer D- over L-arabinose.

(A) CAFÉ assay quantifying flies' intake of D- and L-arabinose across a range of concentrations. **(B)** In 5 minutes flies consume more radioactive ^{32}P-mixed D-arabinose than radioactive ^{32}P-mixed L-arabinose during two-choice tests when either sugar is tested separately against water. **(C)** Single flies monitored by video spend more time on 1M D-arabinose than 1M L- when the two are side-by-side. **(D)** Proboscis extension reflex response to D- and L-arabinose. Flies were given a drop of water, then 10mM D-ara, then 100mM D-ara, then 1M D-ara, water, followed by 10mM L-ara, 100mM L-ara, and 1M L-ara. Alternate flies were given D-ara or L-ara first. **(E)** Flies'

likelihood of detection and consumption of sucrose does not significantly change at concentrations as low as 10mM. **(F)** Flies form better memories with increasing concentrations of sucrose. 10 mM sucrose (the maximum possible contamination given L-arabinose purity of ≥ 99%) produces no significant long-term memory. **(G)** The greater the concentration of D-arabinose, the better the memory score. This suggests that the flies are not eating so much D-arabinose that they become sick, and would otherwise remember if not for the excess sugar. Memory does not improve at lower concentrations. Results with error bars are means ± s.e.m. ns, not significant. *≤0.01, **≤0.001 and *** ≤0.0001.

The neural basis for the difference in behavioral responses to D- and L-arabinose

The neural basis for the behavioral differences towards D- and L- consists of either of two possibilities, not mutually exclusive: the two sugars engage different neural pathways, or they activate the same neural pathway(s) in a different manner. The first points of contact in sugar detection and consumption are the gustatory-receptor-expressing (Gr) neurons that respond to sweet substances. To date, only nine receptors—Gr5a, Gr43a, Gr61a, and Gr64a, b, c, d, e, and f—have been implicated in sweet sugar detection (Dahanukar et al., 2007; Freeman et al., 2014; Jiao et al., 2008; Miyamoto et al., 2013; Yavuz et al., 2014). We therefore used Gr-GAL4 lines—to drive expression of the inward rectifying potassium channel Kir2.1, silencing these sets of Gr-expressing neurons (Baines et al., 2001)—and individual receptor mutants to determine both the neurons and receptors involved in D- and L-arabinose detection, D- vs L- preference, and L-arabinose memory.

Preference for a sugar depends on both detection and discrimination. We began by looking for neurons involved in D- and L-arabinose detection: silencing Gr61a-expressing neurons as well as Gr64f neurons (Gr64f is thought to label all sugar neurons) reduced D-arabinose detection substantially (Figure 7A) (Jiao et al., 2008; Wisotsky et al., 2011). There was no other set of taste neurons whose silencing substantially altered D-arabinose detection. When we turned to the individual receptors, using receptor mutant flies, only removal of Gr64f and the combined removal of Gr64a, Gr64b, and Gr64c showed modest reductions in detection; all other mutants showed normal D-arabinose detection (Figure 7B). In no case did removing a receptor eliminate D-arabinose detection, suggesting a robust system with at least semi-redundant receptors.

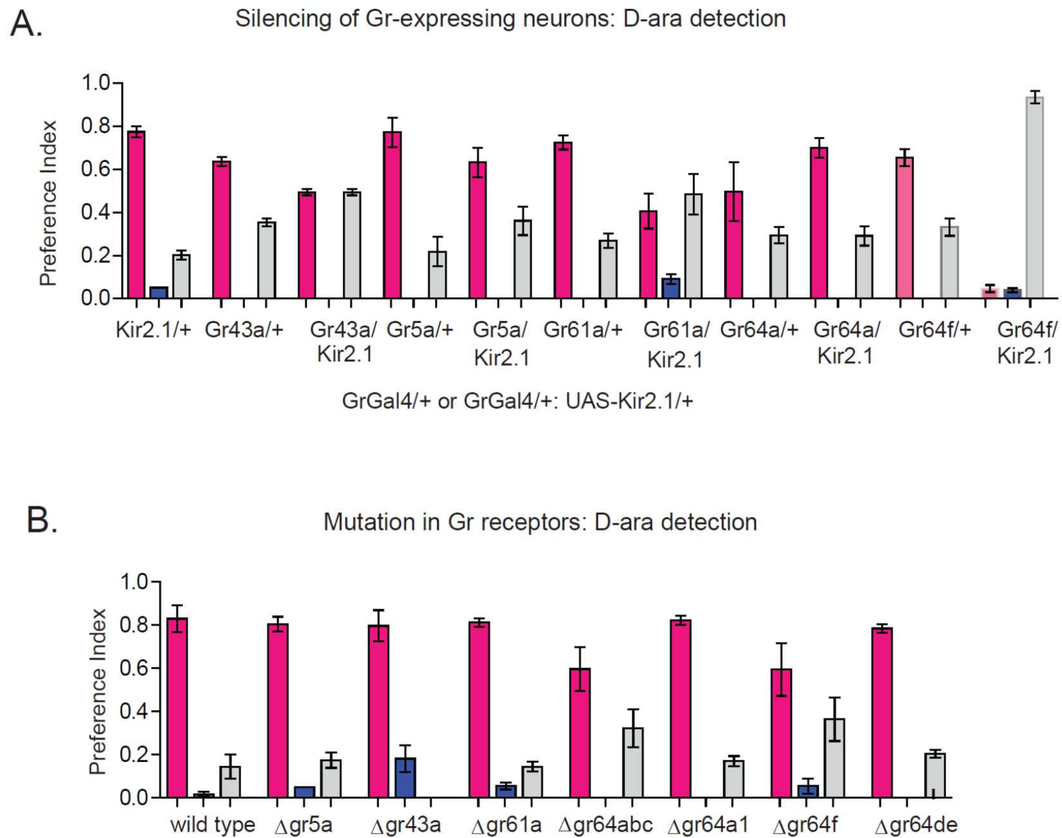


Figure 7. D-arabinose detection does not rely on any single receptor or set of neurons.
(A) Silencing Gr64f-expressing neurons virtually eliminates D-arabinose detection ($p<.0001$); silencing of Gr61a- ($p=.01$) and Gr43a-expressing ($p=.01$) neurons reduces but does not eliminate D-arabinose detection compared to corresponding GAL4-only control. **(B)** Deletion of only Gr64abc and Gr64f reduce D-arabinose detection. Results with error bars are means ± s.e.m. ns, not significant. *≤0.01, **≤0.001 and *** ≤0.0001.

We then looked at which neurons and receptors are involved in L-arabinose detection. Silencing Gr5a-expressing neurons reduced L-arabinose detection by about 80%, while silencing Gr61a-expressing neurons reduced L-arabinose detection by ~50% (Figure 8A). Silencing Gr64f neurons again eliminated sugar detection (Figure 8A). Turning to the receptors again, deletion of Gr43a had a small effect on L-arabinose detection, and deletion of Gr61a resulted in ~40% reduction (Figure 8B). Deletion of Gr64f also reduced L-arabinose detection nearly 50%, and combined deletion of Gr64d and Gr64e had a modest effect (Figure 8B). From these detection assays, it is clear that silencing Gr64f neurons—which likely means silencing all sugar neurons—renders flies incapable of recognizing sugar. Silencing Gr61a neurons

reduces both D- and L-arabinose detection, though does not completely eliminate it. As will be discussed later, Gr61a is expressed in all leg neurons but not many proboscis neurons, meaning that silencing sugar neurons in the leg is probably sufficient to begin to interfere with these sugars' detection. It is clear that L-arabinose detection is more sensitive to individual receptors' removal. While Gr64f was the only single receptor mutant to show even a modest deficit in D-arabinose detection, removal of Gr43a, Gr61a, and Gr64f all showed reductions in L-arabinose detection. In the case of Gr43a, however, the deficit is minimal—the vast majority of flies still have no problem recognizing L-arabinose. Moreover, when the L-arabinose concentration is reduced 10-fold to 100mM, Gr43a-null flies detect L-arabinose as well as wild type flies, arguing that there's no real L-arabinose detection deficit. This is supported by the normal L-arabinose detection seen when all Gr43a-expressing neurons were silenced.

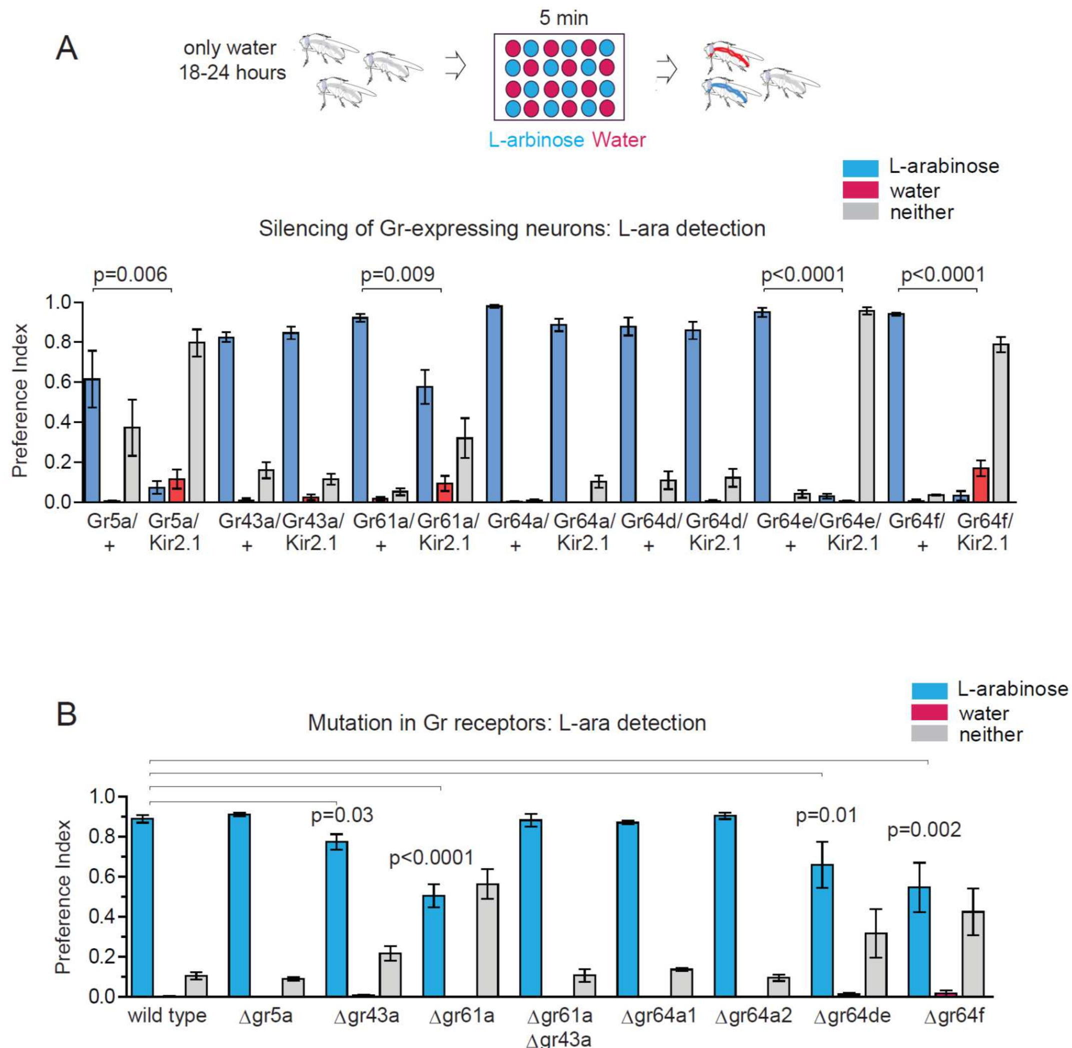


Figure 8. Gr43a and Gr61a neurons and receptors are involved in L-arabinose detection.
(A) Fifty male flies per trial were given water alone for 24-36 hours, then put on a microtiter plate checkered with food-dye-labeled water and 1M L-arabinose. After five minutes, flies were removed and the color visible in the abdomen was scored. *n* = 4 for all silencing and mutant two-choice experiments, 50 flies per *n*. Silencing Gr5a, Gr64e-, and Gr64f-expressing neurons nearly eliminated L-arabinose detection; silencing Gr43a^{GAL4} neurons had no detectable effect on L-arabinose detection. **(B)** Only Gr43a, Gr61a, Gr64d/e, and Gr64f mutants had reduced L-arabinose detection. Results with error bars are means ± s.e.m. ns, not significant. *≤0.01, **≤0.001 and *** ≤0.0001.

We then performed the same assays but with L- and D-arabinose side-by-side, to ask whether any of these neurons or receptors plays a role in the D-arabinose preference. While silencing neurons required for detection would cause a global decline in consumption—as seen when silencing Gr64f neurons—silencing neurons required for discrimination would result in equivalent consumption of D- and L-

arabinose. Upon silencing of Gr5a-, Gr43a-, Gr64a-, or Gr64d-expressing neurons, flies still overwhelmingly preferred D-arabinose (Figure 9A). Silencing Gr64e and Gr64f neurons almost completely abolished consumption of both sugars, consistent with the results for D- and L-arabinose detection above. (Again, silencing Gr64e and Gr64f neurons likely means silencing all sugar-detecting neurons—remove these, and the flies are unable to detect sugar.) The only set of neurons whose silencing impaired D-arabinose preference were the Gr61a neurons: silencing Gr61a neurons reduced D-arabinose consumption while L-arabinose consumption increased (with ~30% flies not eating any sugar), indicating that without Gr61a-expressing neurons flies were having some trouble discriminating between the two sugars (Figure 9A). And again, because Gr61aGAL4 covers probably all sugar-detecting leg neurons but only a handful of proboscis neurons, it seems likely that the legs play a large role in dictating the initial preference.

All individual Gr mutant flies preferred D-arabinose (Figure 9B). About 30% of Δgr64de mutants and nearly 40% Δgr64f mutants did not eat any sugar, but the flies that did still consumed more D than L. Only in the chemosensory transcription factor mutant *poxn*, where no external chemosensory neurons develop (the LSO and VCSO neurons lining the pharynx do develop and are functional), was the discrimination between D- and L-arabinose reduced (Figure 9B) (Awasaki and Kimura, 1997; LeDue et al., 2015).

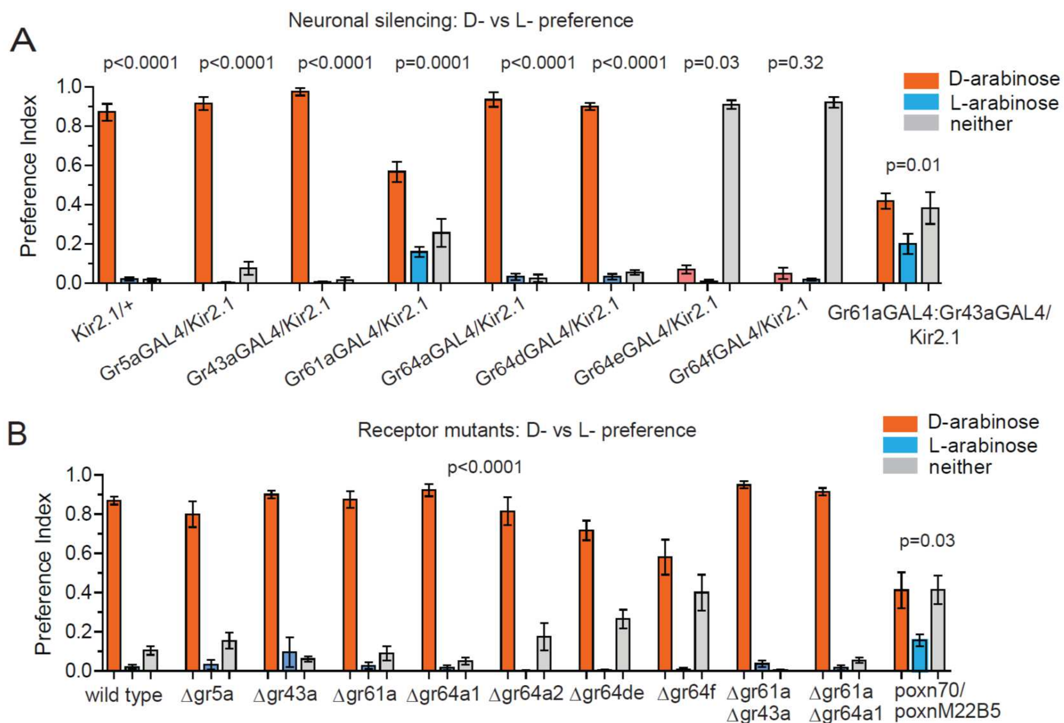


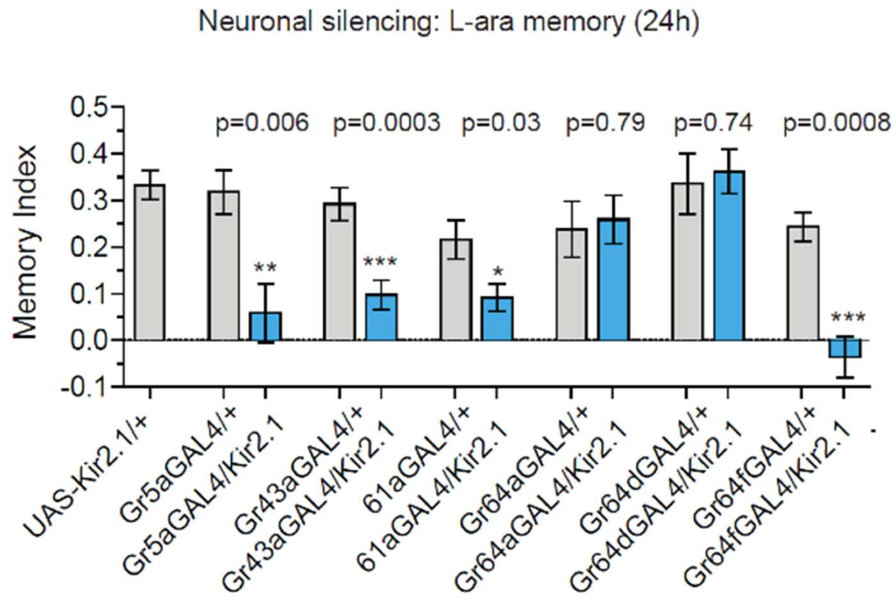
Figure 9. Gr61a-expressing neurons are involved in D- vs L- preference.
(A) Silencing of Gr61a-expressing neurons with Kir2.1 impaired D > L discrimination and preference; silencing Gr64e- and Gr64f-expressing neurons nearly eliminated detection of either sugar. **(B)** No single receptor mutant impaired D > L preference; flies lacking Gr64d/e and Gr64f displayed reductions in the number of flies eating D-arabinose. Results with error bars are means ± s.e.m. ns, not significant. *≤0.01, **≤0.001 and *** ≤0.0001.

Finally, we looked at the same neurons and receptors in the context of long-term memory. For L-arabinose memory, silencing Gr5a, Gr43a, and Gr61a neurons significantly impaired L-arabinose memory—however, Gr5a and Gr61a neurons play important roles in L-arabinose detection (Figure 10A). Therefore these flies could very well be struggling to remember L-arabinose simply because they are having trouble detecting it. Gr43a neurons are unique in that they're dispensable for L-arabinose detection, but silencing them causes dramatic reductions in L-arabinose memory. Because D-arabinose memory is relatively poorer, it is more difficult to clearly identify flies that have problems remembering D-arabinose as opposed to normal variation in the assay, but all lines silenced and all mutants tested had slightly lower D-arabinose memory.

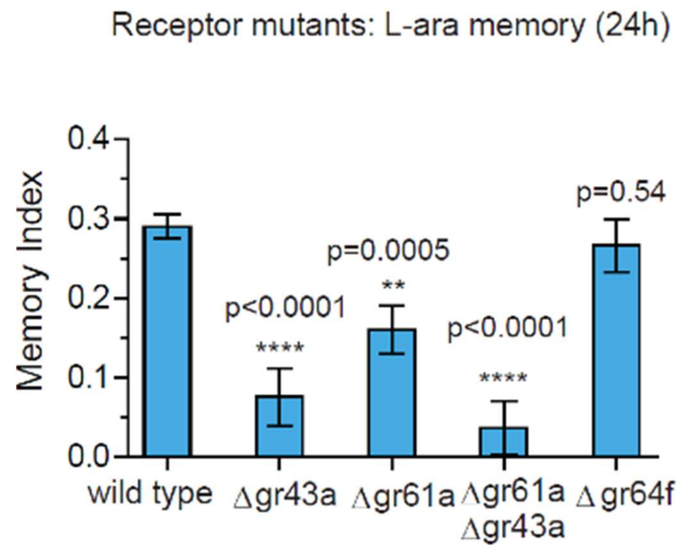
In contrast to D-arabinose preference, silencing of Gr5a-, Gr43a-, and Gr61a-, but not Gr64a- or Gr64d-expressing neurons, significantly impaired L-arabinose memory (Figure 10A). Notably, silencing of Gr43a neurons had no effect on L-arabinose detection (Figure 8A), indicating a more selective role for Gr43a neurons in L-arabinose-mediated memory (Figure 10A). When the receptor mutants were trained with L-arabinose, Δgr43a and Δgr61a showed a significant reduction ($p < 0.01$) in long-term memory at 24h (Figure 10B).

D-arabinose memory was reduced when Gr5a, Gr61a, and Gr64f neurons were silenced; it was similarly reduced when Gr43a, Gr61a, Gr64a, and Gr64abc receptors were removed (Figure 8C).

A.



B.



C.

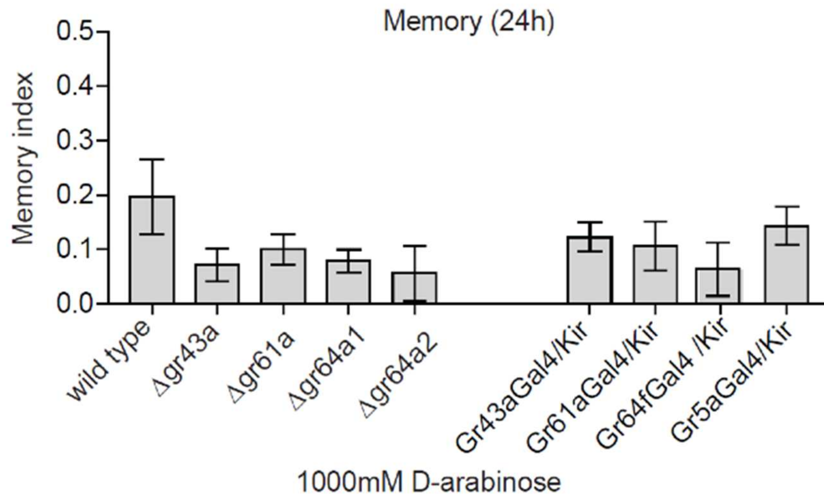


Figure 10. L-arabinose and D-arabinose memory
(A) Silencing Gr43a and Gr61a-expressing neurons impaired L-arabinose memory. Gr64aGAL4 and Gr64dGAL4, whose expression is restricted to LSO and VCSO neurons did not impair L-arabinose memory. Silencing Gr64f- and Gr5a-neurons also reduces L-arabinose memory, but they also impair L-arabinose detection. **(B)** Gr43a and Gr61a receptors are important for L-arabinose memory. **(C)** D-arabinose memory (24hr) is reduced, though in some mutants not quite significantly, by removal of individual receptors. Silencing various sets of neurons impairs D-arabinose memory similar to the impairment in L-arabinose memory. Results with error bars are means ± s.e.m. ns, not significant. *≤0.01, **≤0.001 and *** ≤0.0001.

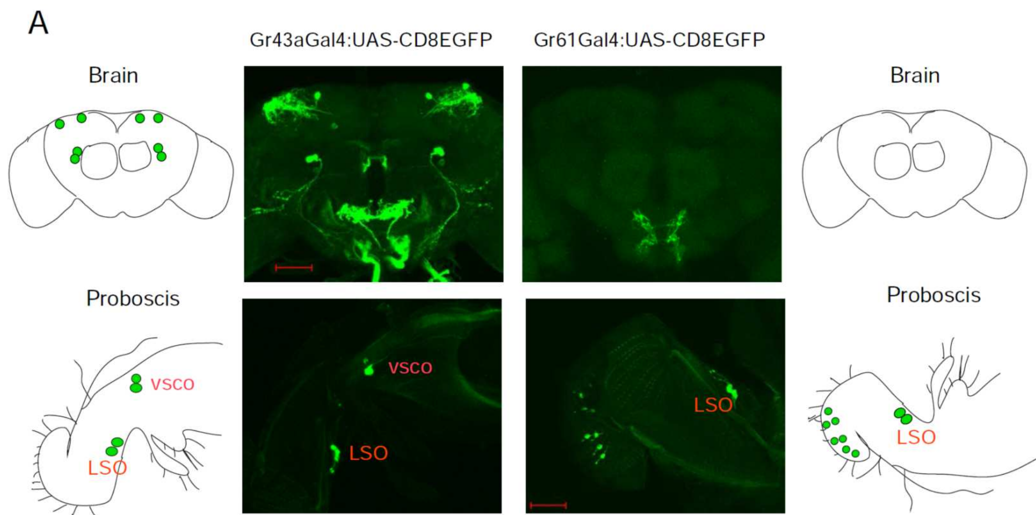
Taken together, these results suggest that preference for D-arabinose over L-arabinose relies heavily on Gr61a-expressing neurons—these are the only set of neurons so far identified that when silenced show a weakened preference for D-arabinose. Gr61a neurons are also involved, as one might expect, in L-arabinose and D-arabinose detection. Gr61a neurons appear to label all sugar-responsive leg neurons, but only a moderate number of proboscis neurons (whereas Gr5a, for example, labels seemingly all proboscis neurons but only a handful of leg neurons). Because silencing Gr61a neurons impairs discrimination between D- and L-arabinose, but silencing Gr5a does not, we think that the leg neurons likely mediate a large portion of the preference for D-arabinose. To test this more precisely we'd like a driver that silences only the leg neurons, but all the leg neurons; unfortunately, the only other drivers that label all leg neurons also label all the proboscis neurons (Gr64e and Gr64f). L-arabinose memory, in contrast, is much more sensitive than L or D detection, and can be interfered with by silencing Gr43a-, Gr61a-, and Gr5a-expressing neurons. Gr61a and Gr5a neurons however, are also involved in L-arabinose

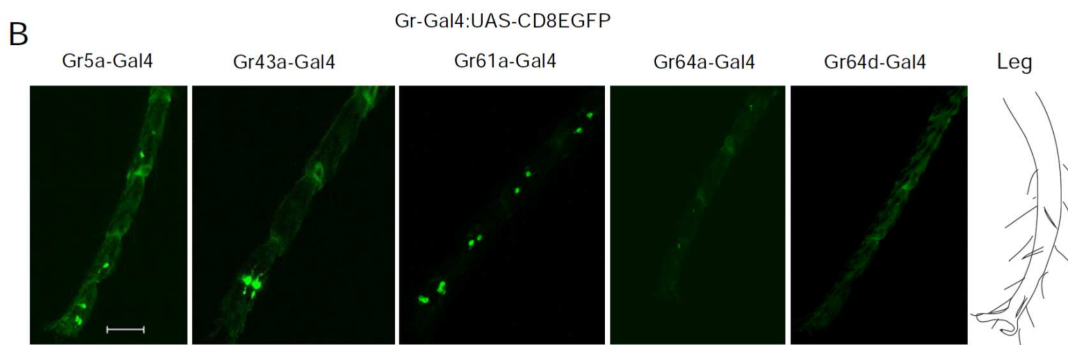
detection, meaning that it's unclear whether the memory deficits seen when silencing Gr5a or Gr61a neurons are true memory deficits or secondary to a problem with detection. At the receptor level, D-arabinose preference is maintained in the absence of any single known sugar receptor (Figures 9B), whereas L-arabinose memory requires Gr43a and Gr61a receptors (Figure 10B). These results, however, do not rule out the possibility that there may be an as-yet unidentified receptor that exclusively mediates D-arabinose preference or that L-arabinose memory may use other receptors, in addition to Gr43a and Gr61a.

Peripheral Gr43a neurons are necessary for L-arabinose memory

In *Drosophila*, gustatory receptors are expressed on the antennae, legs, wings, and labellae, and in the pharynx, gut, and central brain (Joseph and Carlson, 2015). The wide expression pattern, presence of multiple receptors in the same neurons, and different combinations of receptors in different neurons indicate that gustatory-receptor-expressing neurons in various locations may respond quite differently to different sugars (Miyamoto and Amrein, 2014; Thoma et al., 2016). Since Gr61a-, Gr43a-, and Gr5a-expressing neurons are required for L-arabinose memory, and Gr61a- and Gr43a-expressing neurons also contribute to D-arabinose preference, they could all label a common set of critical neurons, or be separately important, perhaps in different parts of the body. As reported by others (Miyamoto et al., 2012; Park and Kwon, 2011), Gr43a^{GAL4} expression is consistently detected in four dorsolateral protocerebrum (DLP) neurons in the central brain, the LSO and VCSO neurons in the proboscis, two f5 neurons in the distal tarsi, and in the proventricular ganglion of the gut (Figure 11A, 11B). Gr61a is expressed in the labellum, tarsi—including the most distal segment—and LSO neurons in the proboscis, but not in the gut or CNS (Figure 11A, 11B). We selectively silenced the central brain DLP neurons using a GAL4:GAL80 combination (Figure 12A) (Miyamoto et al., 2012), or the LSO and VCSO neurons using Gr64aGAL4 and Gr64dGAL4 (Figure 7C). While silencing of all Gr43a neurons impaired L-arabinose memory, silencing of just the DLP, or just the LSO and VCSO Gr43a neurons had no significant effect, suggesting that some combination of Gr43a-expressing neurons that includes the gut and/or tarsal neurons are sufficient for L-arabinose memory. In the f5 neurons of the distal tarsi, Gr43a is coexpressed with Gr5a (Figure 11B) and/or Gr61a, while neither Gr5a nor Gr61a expression can be detected in the gut. Because silencing of Gr61a- and Gr5a-expressing neurons each blocked L-arabinose memory (Figure 7C), it seems likely that the tarsal Gr43a neurons are involved in L-arabinose memory. However, these results do not

rule out the possibility that other Gr43a neurons or other Gr-receptor expressing neurons are also involved in L-arabinose memory.





C

Location	GAL4 expression
Proboscis	Gr5a (widespread; LSO) Gr43a (limited; LSO, VCSO) Gr61a (limited) Gr64a (limited; LSO) Gr64d (widespread; LSO, VCSO) Gr64e (widespread) Gr64f (widespread)

Brain	Gr43a
Gut	Gr43a
Leg	Gr5a (many; 3 in terminal segment, one of which overlaps with Gr43a) Gr43a (only the 2 in last segment) Gr61a (every leg sugar neuron) Gr64a (none) Gr64d (none) Gr64e (every leg sugar neuron) Gr64f (every leg sugar neuron)

Figure 11. Expression patterns of Gr61aGAL4 and Gr43aGAL4.
(A) Gr43a^{GAL4} neurons are observed in the dorsolateral protocerebrum, central brain, proboscis, leg, and gut (not shown). Gr61a-expressing cell bodies are located only in the proboscis and in the leg. **(B)** Distal tarsi expression patterns of Gr5aGAL4 (3 neurons), Gr43a^{GAL4} (2 neurons), Gr61aGAL4 (6 neurons), Gr64aGAL4 (absent), and Gr64dGAL4 (absent). Scale bar 50µm. **(C)** Table of Gr-GAL4 expression. Unfortunately, because of the large numbers of neurons labeled, knowing exact overlap between these drivers is currently problematic, with a couple exceptions mentioned above. Also note that these are not claims about actual expression of the receptor—it's likely a close approximation, but for these drivers we care about the neurons we're manipulating; we don't necessarily make claims about the specific receptors expressed in those neurons.

L-arabinose and D-arabinose differentially activate peripheral Gr43a neurons

To understand how D- and L-arabinose generate different behavioral responses, and because tarsal Gr43a neurons seemed like the best candidates for a small number of neurons critical for L-arabinose memory, we analyzed electrophysiological responses of Gr43a neurons in distal tarsi (f5V) as well as labellar proboscis neurons. In both neuronal populations, D-arabinose consistently generated significantly more spikes over a range of concentrations (Figure 12B-E). Differences in electrophysiological response also corresponded with changes in intracellular calcium levels measured by GCaMP6, a genetically encoded calcium indicator (Chen et al., 2013). In Gr43a-expressing f5 neurons in distal tarsi, the fluorescence peak was higher and reached more rapidly for D-arabinose than L- (Figure 12F). Removal of the Gr43a receptor from these neurons significantly reduced response to L-arabinose but not D-arabinose (Figure 13A), consistent with the idea that D-arabinose responses are robust to the removal of any single receptor. In the proboscis, LSO neurons showed a quicker rise and fall in response to L-arabinose, with a slower but more sustained activation in response to D-arabinose (Figure 13B). However, such differences between D- and L-arabinose-provoked responses were not universal: the average D- and L- arabinose responses of the central brain DLP neurons were similar in both magnitude and shape (Figure 12G). These results indicate that while D- and L- arabinose can activate the same

gustatory neurons to different extents, differential activation depends on properties specific to each neuron.

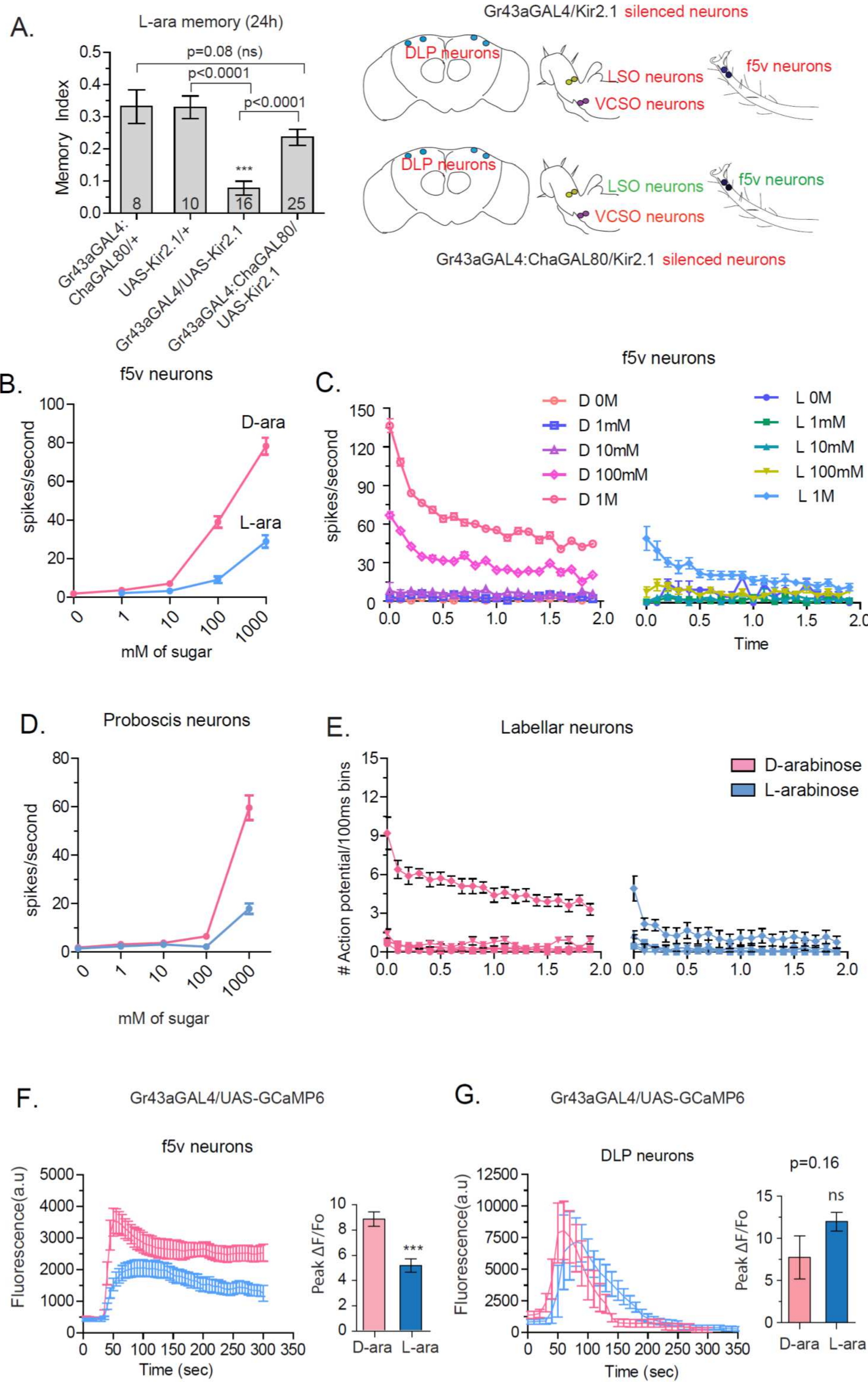
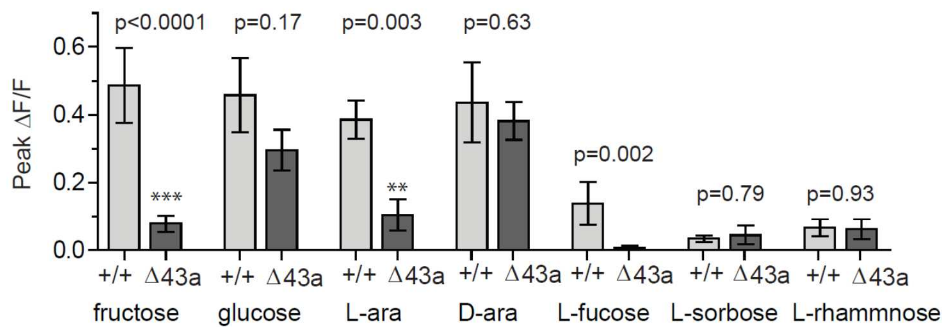

A.
L-ara memory (24h)
p=0.08 (ns)
p<0.0001
p<0.0001
Memory Index
Gr43aGAL4:ChaGAL80/+
UAS-Kir2.1/+
Gr43aGAL4/UAS-Kir2.1
Gr43aGAL4:ChaGAL80/UAS-Kir2.1
Gr43aGAL4/Kir2.1 silenced neurons
DLP neurons
LSO neurons
VCSO neurons
f5v neurons
Gr43aGAL4:ChaGAL80/Kir2.1 silenced neurons
B.
f5v neurons
spikes/second
D-ara
L-ara
mM of sugar
C.
f5v neurons
D 0M
D 1mM
D 10mM
D 100mM
D 1M
L 0M
L 1mM
L 10mM
L 100mM
L 1M
Time
D.
Proboscis neurons
spikes/second
mM of sugar
E.
Labellar neurons
Action potential/100ms bins
D-arabinose
L-arabinose
F.
Gr43aGAL4/UAS-GCaMP6
f5v neurons
Fluorescence(a.u)
Time (sec)
Peak ΔF/Fo
D-ara
L-ara
G.
Gr43aGAL4/UAS-GCaMP6
DLP neurons
p=0.16
ns
Fluorescence(a.u)
Time (sec)
Peak ΔF/Fo
D-ara
L-ara

Figure 12. Peripheral Gr43a neurons are critical for L-arabinose memory and are activated differentially by D- and L-arabinose.

(A) Silencing only the dorsal protocerebral (DLP) and VCSO neurons does not impair L-arabinose memory. Left panel: memory score in various genetic background. Right panel, top: in Gr43a^{GAL4}/Kir2.1 flies, all indicated neurons are silenced (including proventricular neurons, not pictured). Bottom: in Gr43a^{GAL4}:ChaGal80/UAS-Kir2.1 flies, only the neurons indicated in red are silenced. **(B)** Spikes per second of recorded labellar f5V tarsal neuron in response to D- or L-arabinose at various concentrations. **(C)** Spikes per second binned by 100ms over the first two seconds of response. In all cases, D-arabinose evoked many more spikes/second than L-arabinose, and higher concentrations evoke more spikes than lower concentrations. **(D)** Spikes per second electrophysiological response of proboscis neurons to increasing concentrations of D- and L-arabinose. **(E)** Action potentials per 100ms for various concentrations of D- and L-arabinose—for both, the highest curve is at 1M. **(F)** The fluorescence response of GCaMP6med-expressing Gr43a^{GAL4} distal tarsi neurons to 1M D- and L-arabinose. **(G)** Imaging of calcium responses of dorsal protocerebral neurons. The left panel shows curve-aligned and normalized fluorescence to compare the shape of the responses, and the right panel plots peak magnitudes as $\Delta F/F_o$. Results with error bars are means ± s.e.m. ns, not significant. *≤0.01, **≤0.001 and *** ≤0.0001.

A.



B.

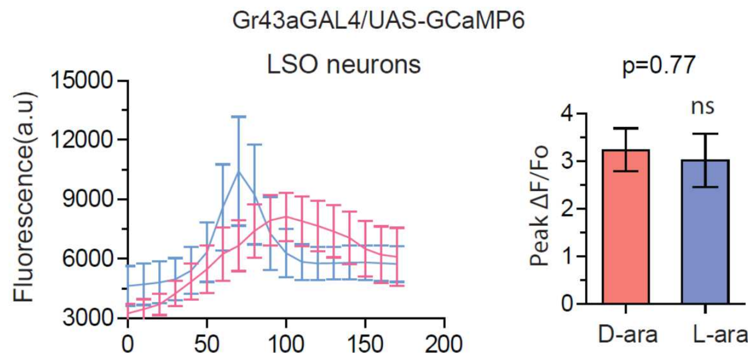


Figure 13. Further activity of Gr43a neurons in response to D- and L-arabinose
(A) The normalized calcium imaging responses of $Gr43a^{GAL4}$ neurons in the distal tarsi. Removal of Gr43a impairs fructose and L-arabinose activation of $Gr43a^{GAL4}$ neurons. **(B)** Imaging of calcium responses of LSO neurons in proboscis. The left panel shows curve-aligned and normalized fluorescence to compare the shape of the responses, and the right panel plots peak magnitudes as $\Delta F/F_o$. Results with error bars are means ± s.e.m. ns, not significant. *≤0.01, **≤0.001 and *** ≤0.0001.

Activation of Gr43a neurons in the absence of sugar can form associative memory

$Gr43a^{GAL4}$ labeled the most restricted set of neurons that was critical for L-arabinose memory. We therefore sought to determine whether they also represented the minimum set of gustatory neurons sufficient to drive appetitive long-term memory formation, and, in turn, whether gustatory neuron activation on its own could drive formation of appetitive memories. To this end we asked whether activating $Gr43a^{GAL4}$ neurons in the memory paradigm—in the absence of sugar—could generate an associative-appetitive memory (Figure 14A). In essence, this would mean causing the flies to remember eating a sugar that was never actually present. dTrpA1, a temperature-sensitive cation channel, causes continuous activation of neurons at temperatures above 26ºC (Hamada et al., 2008). Activation of $Gr43a^{GAL4}$ neurons by dTrpA1, however, failed to substitute for the sugar reward (Figure 14B), although similar activation of a subset of dopaminergic neurons (R58E02-GAL4/+; dTrpA1/+) produced long-term appetitive memory as reported by others (Liu et al., 2012a) (Figure 15A). These results suggested either that activation of $Gr43a^{GAL4}$ neurons is necessary but not sufficient for L-arabinose memory, or that dTrpA1 does not approximate the activation required to produce long-term memory. Consistent with the latter possibility, activation of Gr43a neurons with the red-shifted channelrhodopsin variant ReaChR—a light-gated cation channel that depolarizes neurons in response to red light (Inagaki et al., 2014)—did generate associative memory (Figure 14C). When flies expressing ReaChR in $Gr43a^{GAL4}$ neurons were exposed to one odor without the light, and a second odor in the presence of red light, they showed a greater preference for the light-associated odor 24 hours later (Figure 14C). Since in addition to $Gr43a^{GAL4}$ silencing, Gr5a- and Gr61a-neuron silencing also impaired L-arabinose memory, we also tested the ability of Gr61a and Gr5a neuron activation to generate memory. Only the activation of Gr5a-expressing neurons resulted in robust long-term memory; Gr61aGAL4 activation showed no significant memory above controls (Figure 14D). Activation of Gr64a-expressing neurons, that is, activation of just the LSO and VCSO proboscis neurons, did not produce significant long-term memory (Figure 14D). (Optogenetic activation of additional GrGAL4 drivers is seen in Figure 16B.)

Finally, starvation is an important regulator of memory strength in the associative-appetitive paradigm—the hungrier flies are, the better the memories they form (Colomb et al., 2009; Krashes et al., 2009). Starvation also influenced the memory strength following Gr43a-neuron activation: the same pulsated light activation produced memory in starved but not fed flies (Figure 14E). Taken together, these results suggest that activation of a subset of Gr43a-expressing neurons is sufficient to generate long-lasting associative memory. These observations further suggest that activation of the same neurons by different methods, perhaps leading to different activation level or pattern, can give rise to substantially different behavioral outcomes, consistent with other reports (Figure 14F) (Clark et al., 2013; Seeger-Armbruster et al., 2015).

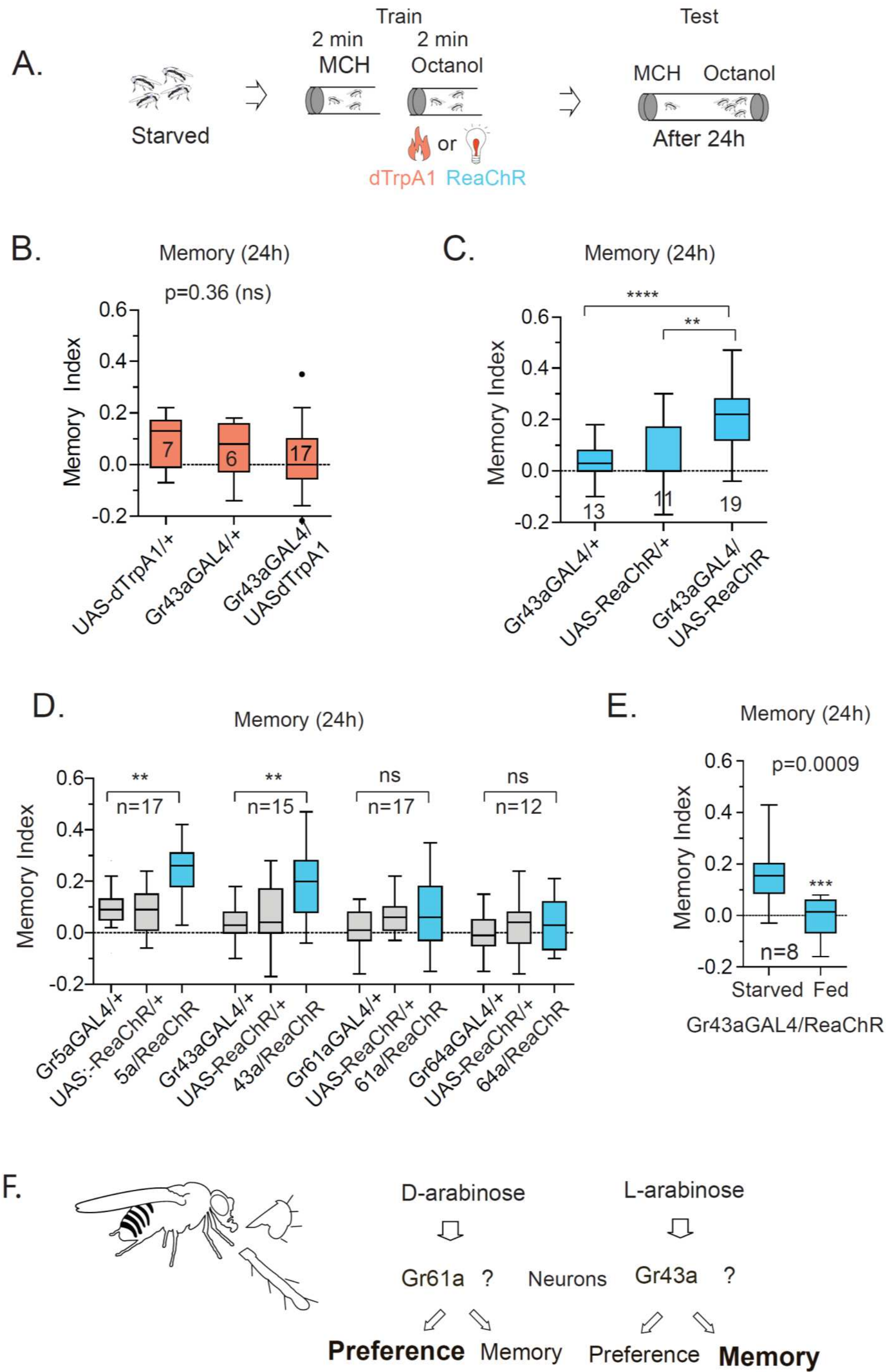


Figure 14. Activation of Gr43a neurons is sufficient for appetitive-associative memory. (A) Schematic of heat and light-activated associative olfactory training. Please see Methods and Figure 4—figure supplement 1C for detail. **(B)** Continuous activation of Gr43aGAL4 neurons by dTrpA1 (at 31 ºC) does not induce long-term memory. **(C)** A

20 Hz, 15 msec pulse-width activation for 2.5 seconds, repeated every 20 seconds, induces long-term memory in flies expressing ReaChR in Gr43a^{GAL4} neurons; genetic controls do not show significant memory, and the same amount of light using the same pulse-width but distributed uniformly over the two-minutes induces poorer memory (red). Schematics of light patterns are not to scale. **(D)** Optogenetic activation of Gr43a- and Gr5a-expressing neurons leads to substantial 24-hour memory; activation of Gr61a- or Gr64a-expressing neurons does not. Activation pattern is the same used in the first three genotypes of Figure 5B. **(E)**: Optogenetic activation of Gr43a^{GAL4} neurons induces memory only in hungry flies, not flies fed ad libitum. **(F)** Model for D-arabinose preference and L-arabinose memory: D-arabinose activates Gr43a-Gr61a and other neurons in a manner that leads to strong preference but weaker memory; L-arabinose activates Gr43a-Gr61a-and/or Gr43a-Gr5a coexpressing neurons in a manner that leads to weaker preference but stronger memory. Results with error bars are means ± s.e.m. ns, not significant. *≤0.01, **≤0.001 and *** ≤0.0001.

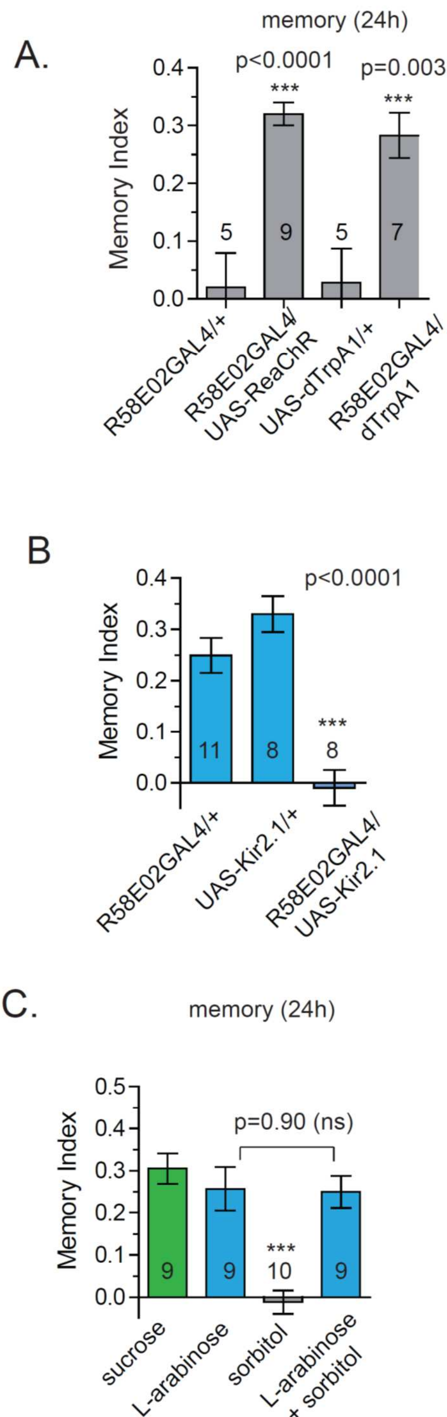


Figure 15. dTrpA1 activation of Gr43a neurons does not produce long-term memory.
(A) Activation of R58E02GAL4 dopaminergic neurons, either with light (ReaChR; 1 Hz, 500 msec pulse-width for 20 seconds, repeated three times over two minutes, or 10 Hz, 10 msec pulse-width, repeated continuously for two minutes, data combined) or with heat (dTrpA1; 31 ºC for two minutes) is sufficient to generate robust 24hr memory. **(B)** Silencing of R58E02GAL4 neurons eliminates L-arabinose memory. **(C)** Supplementing L-arabinose with 1M

sorbitol does not increase L-arabinose memory. Results with error bars are means ± s.e.m. ns, not significant. *≤0.01, **≤0.001 and *** ≤0.0001.

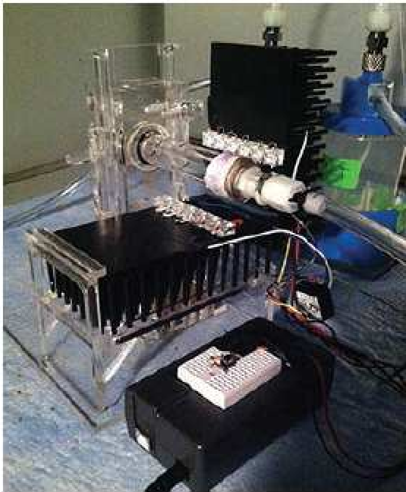
Optogenetic activation

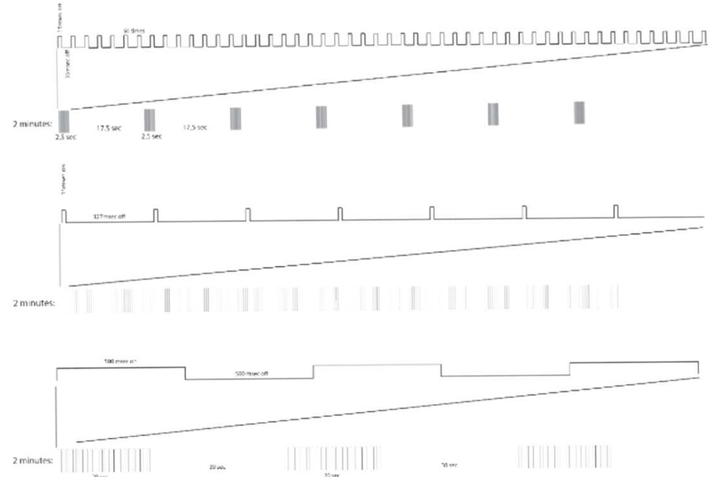
As detailed above, our initial efforts to determine whether activating Gr43a or Gr61a neurons (neurons required for L-arabinose memory) would be able to generate memory in the absence of sugar began with dTrpA1, a *Drosophila* heat-activated cation channel. Because the flies with the dTrpA1 channel showed no evidence of memory formation, we didn't know whether dTrpA1 wasn't working, the set-up wasn't able to activate dTrpA1, or whether we were activating these neurons just fine but that activating only Gr43a neurons would never generate false memories of sugar. To determine which of these possibilities we were facing, we used a dopaminergic GAL4 (R58E02GAL4) that another group had used to generate memory by dTrpA1 (Liu et al., 2012a). If heating these flies in the context of the memory paradigm generated 24-hour memory, we would know the set-up was working and our problem was something more specific to Gr neurons. And this was exactly what we saw: after twenty-four hours, we saw good memory in the R58E02GAL4 flies (Figure 15A). At this point it wasn't clear whether Gr43a neuron activation would never produce memory, or whether some kind of activation might be able to, just not using dTrpA1.

The obvious difference between the two methods of activation, apart from heat versus light, is that light can be turned on and off instantaneously, while most practical methods of heat activation have longer lags between heat onset and offset. The first indication that this could matter in sugar-sensing neurons was an experiment comparing the proboscis extension reflex (PER) between dTrpA1 and ReaChR activation. (Many thanks to Liying for showing me the paper.) When flies taste sugar (or think they're tasting sugar) they instinctively extend their mouth as if to feed. The same reflex is prompted if channelrhodopsin is used to activate sugar-sensing neurons. One light pulse of just 100msec was able to provoke the proboscis response. However, in no flies expressing dTrpA1 was heat activation able to prompt proboscis extension (Inagaki et al., 2014). Whether due to the relative delay in heating kinetics, or some other difference in activation pattern, dTrpA1 could not prompt this sensory-neuron-driven behavior, while ReaChR could. We therefore decided it was worth trying ReaChR activation of Gr43a neurons in the memory paradigm. Unlike heat activation, however, light pulse activation presents innumerable choices about activation pattern (Figure 16A): Continuous light for two minutes? Pulsed light? How long should the pulses be? How far apart should they be spaced? After three trials each of several patterns that didn't seem to generate memory, a pattern of 15msec light on, 35msec light off,

repeated 50 times (which takes ~2.5 second), with that bursting pattern repeated once every twenty seconds, appeared to produce fairly substantial memory (Figure 16B). One reminder: the light pattern is not the exact neuronal firing pattern—the light of course causes the firing, but given different tissue densities, channelrhodopsin expression, and neuron properties, the same light pattern can in different contexts give rise to variable neuronal firing patterns. Moreover, adjustments to the light pattern will not necessarily provoke firing pattern changes in a linear or predictable manner.

A.





B.

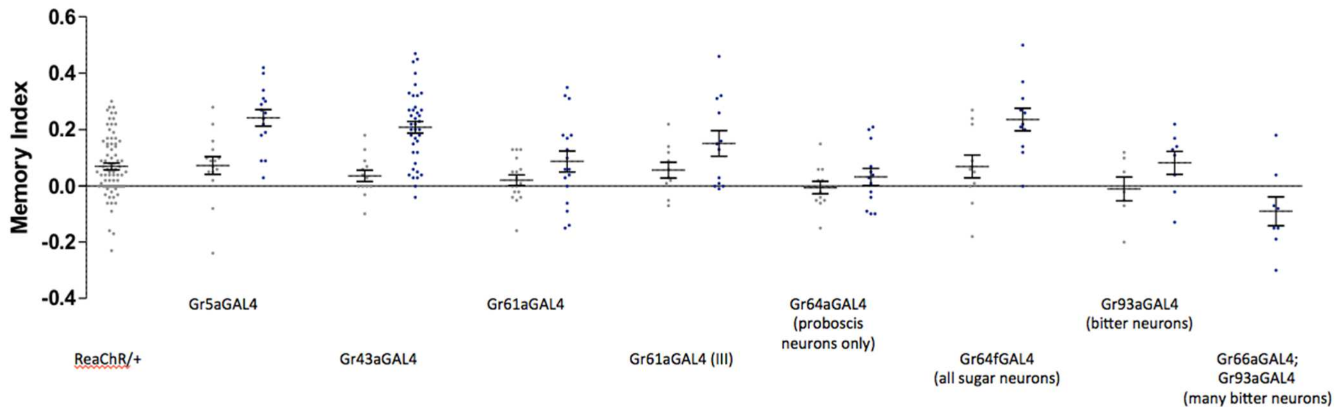


Figure 16: Optogenetic activation of GrGAL4 neurons to create false appetitive memories

(A) Left: The optogenetic rig with two panels of LEDs aligned with the training chamber. Right: Three optogenetic activation protocols; the top protocol was used for all GrGAL4 data. **(B)** Optogenetic activation of Gr5aGAL4, Gr43aGAL4, and Gr64fGAL4 neurons produce significant 24 hour memory; Gr61aGAL4 on the third chromosome but not the second produces marginally significant memory. Activation of Gr93aGAL4 or Gr66aGAL4:Gr93aGAL4 bitter neurons produces no memory. ReaChR/+ controls are combined on left; GrGAL4/+ controls are the first (grey) group for each, while GrGAL4/ReaChR flies are in blue.

One requirement of channelrhodopsin is the presence of all-*trans* retinal, the light-responsive chromophore whose absorbance of a photon causes the channel to open. Though the retinal requirement was initially thought to be a significant hurdle to ectopic expression of channelrhodopsins, it was found that vertebrate's endogenous retinal levels were sufficient, and that invertebrates could be fed retinal (Deisseroth, 2015). Our initial experiments had not used retinal, but we found that adding 40uM all-*trans* retinal to the water twenty-four hours before training improved ReaChR activated memory to nearly 80% of that seen with L-arabinose. The remaining difference between our optogenetically induced memory and the sugar memory could be due to several factors, including additional neurons we aren't activating, a natural activation pattern that we aren't perfectly approximating, or some other effect that actually consuming the sugar has. Adding retinal also improved several activation patterns that without retinal had appeared to produce no memory.

The question of why activating R58E02 (dopamine) neurons by either method—dTrpA1a and ReaChR—is sufficient to generate long-term memory, but only ReaChR can activate Gr43a neurons in a manner able to generate long-term memory is a puzzle. The most theoretically attractive explanation is that these sensory neurons have more stringent requirements for the manner of activation able to engage downstream circuitry. The proboscis extension experiments using dTrpA1 support this idea: electrophysiological recordings of Gr5a neurons expressing dTrpA1 do fire in response to heat, but that firing does not lead to proboscis extension (Inagaki et al., 2014). This ties into one of the most curious observations of the optogenetics era: the importance of the activation pattern.

From the first optogenetic experiments in living animals, it became clear that neural circuit-driven behaviors could be driven by light patterns (and therefore by firing patterns) that did not match those found in vivo (Miesenbock, 2009). As has been pointed out, we do not have much data on what naturalistic neuronal activation patterns look like—electrophysiological experiments often require sedating the animal, presenting controlled stimuli, and recording from a tiny fraction of the neuronal ensemble involved in the sensation or behavior. The ideal experiment, of course, would be to electrophysiologically monitor thousands or even hundreds of thousands of neurons while the animal is behaving normally in a natural setting. To determine which of those neurons and firing patterns is

important, we would then need a system capable of simultaneously playing back each neuron's activity pattern to each individual neuron, altering the patterns as necessary to determine causative features. Obviously, we're not currently in a position to do that. And given our limited knowledge of naturalistic firing patterns, it also seems unlikely that we have been able to guess the right patterns, in the right neurons, to cause flies to fly and court, fish to swim, mice to wake up, and cortical neurons to synchronize (Miesenbock, 2009). It instead seems more likely that we do not need to precisely mimic the naturalistic activation pattern. This 'unreasonable effectiveness' of optogenetics has led to the suggestion that the neurons examined so far operate according to attractor dynamics: activate these neurons in something kind of close to a somewhat reasonable pattern and the system will funnel that activation into an interpretable signal (Miesenbock, 2009). But if that's true, the question remains: different neurons often fire with different frequencies—if those frequencies turn out not to matter for neural communication, what are the variety of natural frequencies for?

DISCUSSION

In the course of exploring both the immediate appeal of various natural sugars and their ability to generate long-term associative memories, we fortuitously discovered that these two processes are separable. A specific illustration of this phenomenon is seen with a pair of chemical isomers, D- and L-arabinose: flies greatly prefer D-arabinose to L, but more reliably form long-term memories of L-arabinose (a sugar released from pectin as fruit ripens) (Ahmed and Labavitch, 1980; Dick and Labavitch, 1989). We have also begun to explore the way in which an animal assesses whether an experience that is rewarding in the moment is also of long-term relevance. We find that D- and L-arabinose differentially activate many of the same gustatory-receptor-expressing neurons, including those critical for L-arabinose memory. Many studies have focused on characterizing the higher-order systems, particularly neuromodulatory systems such as dopamine (Berry et al., 2012; Huetteroth et al., 2015; Liu et al., 2012a; Musso et al., 2015; Schwaerzel et al., 2003; Yamagata et al., 2015), octopamine (Burke et al., 2012; Schwaerzel et al., 2003), neuropeptide F (Krashes et al., 2009), and mushroom body neurons (Aso et al., 2014; Kirkhart and Scott, 2015; Vogt et al., 2014) underlying long-term sugar reward memory in *Drosophila*. How various sugars differentially engage the higher-order reward system, however, remains unclear. We find that peripheral gustatory neurons at least partially mediate this discrimination process, contributing to rapid evaluation and memory of potential rewards.

Attributes of sugars important for long-term memory

The caloric value of a sugar is known to be an important determinant of long-term appetitive memory (Burke and Waddell, 2011; Fujita and Tanimura, 2011; Musso et al., 2015), implying that flies quickly metabolize the sugar and that caloric evaluation somehow provides cues necessary to elicit long-term memory. We find that sugars with no caloric value can also produce long-term appetitive memory. One obvious possibility is that memories of sweet nutritious sugars are distinct from memories of sweet non-nutritious sugars. However this seems so far not to be the case: a group of higher-order dopaminergic neurons (R58E02GAL4) necessary for long-term memory of nutritious sucrose (Liu et al., 2012a) are also necessary for memory of non-nutritious L-arabinose (Figure 15B). Similarly, addition of sorbitol, a tasteless but nutritious sugar, enhances the memory of non-nutritious sugars like xylose and D-arabinose, but does not enhance memory of nutritious sugars (Burke and Waddell, 2011). Adding sorbitol to L-arabinose had no additive effect on long-term memory (Figure 15C). It therefore appears that L-arabinose memory uses at least some of the same neural circuitry as memories of nutritious sugars.

Whether memory of L-arabinose, a non-nutritious sugar, is an exception or represents a more general phenomenon is unclear since we have tested only a limited number of sugars in a particular behavioral paradigm. However, in addition to L-arabinose, L-fucose can also produce memory, and both are components of the pectin in many fruits' cell walls (Ahmed and Labavitch, 1980; Dick and Labavitch, 1989). It is therefore possible that these sugars may signal some specific attributes of ripening fruit—ripening is accompanied by breakdown of the pectin in fruits' cell walls—although neither of these sugars are present in fruits anywhere near the concentrations (1M) used in memory assays. Nonetheless, these observations suggest that flies can quickly assess salient features of sugars—perhaps a leading indicator of nutritional value—without the sugar's metabolic breakdown (Figure 3E). This approach to memory formation allows flies to quickly recognize and remember potential foods using specific cues, a time advantage that could be vital in natural contexts.

Do insects distinguish structurally similar sugars? The taste modality of insects, particularly *Drosophila,* is reported to have limited discriminatory power, and is thought to be based primarily on the intensity of the stimuli as opposed to the chemical nature of the sugar (Masek and Scott, 2010). We do find, apart from flies' differential preference for various sugars at equal concentrations, that for immediate and short-term memory this is largely true. However, we did not observe any obvious correlation between immediate behavior and long-term memories: flies immediate preference is L-fucose > D-arabinose > L-arabinose > L-sorbose; for short-term memory, D-arabinose = L-sorbose ≥ L-arabinose = L-fucose; but in

order of long-term memory score, L-arabinose ≥ L-fucose ≥ D-arabinose = L-sorbose (Figure 3C-E). These results are consistent with the idea that while short-term responses are guided by palatability, long-term behavioral responses are guided by additional attributes of the sugars. It is not yet clear why D-arabinose is a less effective stimulus. Since D- and L-arabinose are both sweet, they may be appealing in slightly different ways, or perhaps D-arabinose carries a negative value that over time reduces the positive association formed initially (or dampens the behavioral output).

Role of gustatory receptors in long-term appetitive memory

The gustatory receptors Gr5a, Gr43a, Gr61a, and Gr64a-f have been implicated in sugar detection (Dahanukar et al., 2007; Dunipace et al., 2001; Freeman and Dahanukar, 2015; Fujii et al., 2015; Jiao et al., 2007; Jiao et al., 2008; Joseph and Carlson, 2015; Montell, 2009; Scott et al., 2001). Although exactly which Gr receptors are responsible for detecting which sugars remains somewhat controversial, two features of sweet-sensing Gr receptors are generally agreed upon: first, different gustatory neurons express a number of Gr receptors in unique combinations; second, more than one receptor is typically involved in detecting a sugar (Fujii et al., 2015). However, the physiological consequences of this combinatorial expression of semi-redundant gustatory receptors remain uncertain. This study supports the idea that gustatory neurons in different locations, expressing unique combinations of receptors, are responsible for discriminating chemically similar sugars and eliciting different behavioral responses (Thoma et al., 2016). Consistent with this, previous studies suggested that Gr43a neurons in the central brain monitor hemolymph fructose levels and modulate feeding behavior (Miyamoto et al., 2012), while we find that these neurons are dispensable for L-arabinose memory, and that peripheral Gr43a-neurons are likely sufficient to signal the presence of a rewarding sugar and generate associative memories. These differences likely arise from the locations of these neurons, and perhaps certain differentially expressed receptors, the presence or absence of various co-receptors, and the second-order neurons to which these sensory neurons project. It seems likely that several of these possibilities exist together—it's intuitively appealing to imagine each neuron expressing its own ensemble of receptors, which would lead to relatively distinct firing behavior that the second-order neurons would either transmit in some form or begin to decode. But far from being able to clarify this picture, it still remains unclear even how many Gr43a-, Gr61a- and Gr5a-expressing neurons in the tarsi are sufficient for L-arabinose memory. (Of course, if the system required individual neurons to fire in their distinct manner, optogenetic activation of 26 neurons in the exact same way would likely not lead to memory formation; the stringency of firing requirements is a topic that will take some work to resolve.) We find that activation of Gr43a-expressing

neurons by ReaChR but not dTrpA1 is able to generate appetitive memory, while artificially activating a subset of dopaminergic neurons (R58E02GAL4) by heat (dTrpA1) or light (ReaChR) both led to long-term memory (Figure 15A). How a difference in activity at the sensory level is conveyed to higher-order neurons, and how that difference is interpreted by the higher-order neurons, remains unclear.

How can structurally similar sugars generate differential activation? Although these sugars bind to some of the same receptors, the relative affinity of the receptors likely varies. In this regard the fly sweet taste system may be similar to that of the mammalian system, where a single heteromeric receptor (T1R2 and T1R3) is responsible for detecting a large number of sweet substances, with multiple and discrete ligand-binding sites in each receptor responsible for generating diverse responses (Yarmolinsky et al., 2009). We suspect that the differential engagement of multiple gustatory receptors leads similar chemicals to generate differential activation of the same neurons, and that differential activation and different ensembles of activated neurons allow higher-order decoding of a sugar's relevant features. We speculate that, at least in *Drosophila*, evaluation of a sugar's long-term salience may be encoded in the activation pattern of subsets of gustatory neurons, which allows rapid evaluation and remembering of nutritious food in complex environments.

Reward perception

The role that rewards play in learned behaviors has been the subject of intense focus for decades. Rewards can, most simply, increase the frequency or intensity of behaviors that occur naturally—this process is known as reinforcement (Berridge, 2007). Rewards can also lead to instrumental conditioning, where the animal learns novel behaviors that help it acquire a reward. This latter phenomenon is sometimes called goal-directed control, in the sense that the new behavior is performed in order to achieve the goal, typically some sort of reward (O'Doherty et al., 2017). These classes of learned behaviors have been well-studied, and have generated ideas like the dopamine prediction error hypothesis (the idea that the greater the difference between the expected reward and actually received reward, the more dopamine is released—that dopamine encodes the difference between expected and actual outcomes). But a less well-known idea, though one that provides a nicely coherent explanation of our D- and L-arabinose results, seeks to be more precise about a reward's properties, and the animal's perception of the reward. This incentive salience theory argues that typical rewards impart two simultaneously reactions: 'liking' and 'wanting' (Berridge, 2007; Berridge and Robinson, 2003; Berridge et al., 2009). Liking is the momentary pleasure that occurs while experiencing the reward; wanting is the drive or motivation to experience the reward, even when the reward is not currently present. In the words of

Berridge and Robinson, the originators of this theory, ‘wanting’ is the response to rewards that transforms mere sensory information about rewards and cues into “attractive, desired, riveting incentives” and “emerged early in evolution as an elementary form of stimulus-guided goal direction, to mediate pursuit of a few . . . unconditioned stimuli” (Berridge and Robinson, 2003). The incentive salience theory says that this ‘wanting’ aspect can be transferred from the actual reward (i.e., unconditioned stimulus) to the cue (i.e., conditioned stimulus)—and it is this transfer that we call learning. Of course, in most cases rewards that are ‘liked’ are also ‘wanted’, and in conventional formulations they are considered effectively identical. But work in mice and humans suggests that ‘wanting’ and ‘liking’ are in fact dissociable, in support of this framework. To begin with, ‘wanting’ is mediated by the dopaminergic system, while ‘liking’ is mediated by the opioid and endocannabinoid systems. Mice deficient in dopamine experience no obvious urge to experience rewards or even to eat, but when directly fed sweet food they display the stereotyped hedonic behavior identical to normal mice given sweet food (Berridge and Robinson, 2016). Conversely, activating certain dopaminergic neurons can cause ‘wanting’ behavior, while either leaving ‘liking’ behavior unchanged or sometimes even reducing it, as when dopaminergic neuron stimulation increases feeding behavior but animals display more facial aversion.

In our case, the observation that two similar sugars—D- and L-arabinose—generate strikingly different levels of behavioral responses can perhaps be best understood using this incentive salience framework (Berridge and Robinson, 2003). It is possible that D-arabinose more strongly engages the ‘liking’ circuitry and L-arabinose the ‘wanting’. (This is not to say that L-arabinose is not also liked compared to no sugar, or that flies aren’t motivated by D-arabinose, just that these sugars differentially engage these processes.) This would explain why, when both sugars are accessible, flies will eat more of D-arabinose than L-, but show greater motivation to return to the L-arabinose-associated odor than the D-arabinose-associated odor. If this interpretation is correct, this work identifies higher cognitive processes in fruit flies, a system eminently tractable for circuit exploration. With several orders of magnitude fewer neurons than humans or mice, fruit flies can provide an example of how a more limited nervous system can carry out these sophisticated processes. While work on higher-order structures has identified different circuitries involved in shorter-term memory and longer-term memory, distinct circuitries for ‘liking’ and ‘wanting’ have yet to be identified in Drosophila.

That two very similar sugars can trigger separately varying levels of liking and wanting, however, suggests to us that there are in fact two different circuitries (at least partially non-overlapping). But in addition to the circuitries that underlie liking and wanting, another important and fascinating question is how differential activation of gustatory neurons can lead to differential engagement of these circuitries.

(Yamagata et al., 2015). Perhaps activity-dependent tracing upon D- and L-arabinose ingestion can identify these respective liking and wanting circuitries. For *Drosophila*, the incentive to remember L-arabinose is perhaps owing to the fact that it can inform about a specific attribute of food, such as the ripening status of fruit. Moreover, work in humans suggests that although 'liking' and 'wanting' both represent a positive reward, they utilize distinct neural processing (Schultz, 2006; Wise, 2002). Our observations with D- and L-arabinose now provide an opportunity to explore the neural basis of 'liking' and 'wanting', and how these reward percepts strengthen memory in the accessible nervous system of *Drosophila*.

PART II. ORB2'S CONFORMATIONAL SWITCH AS A BIOCHEMICAL SUBSTRATE OF MEMORY

The above work on sugar detection and the sensory circuits involved in long-term memory examined the first steps in the process that generates an associative memory. But to create a memory, neuronal activity must ultimately lead to appropriately persistent molecular changes in a select number of synapses. The following projects focus on two proteins that are important in the formation and maintenance of these neuronal changes. In particular, Orb2's conformational switch in response to neuronal activity, and its ability to sustain that altered conformation over time, could very well be a molecular trace of memory itself.

Synapse-specific molecular changes and local protein synthesis

As described in the introduction, neuronal activation that leads to long-term changes in synaptic strength necessarily initiates a signaling cascade that leads to new transcription and translation. These new mRNAs and proteins are then transported from the cell body back to the synapse to maintain the activity-induced structural modifications. But given that the typical human neuron has roughly 7,000 synapses (Drachman, 2005), how can a neuron selectively transport mRNA and protein, beginning in the cell body, all the way out to the one or two synapses that were previously activated? The current thinking, backed by evidence from isolated neuronal cultures, suggests that the neuron actually doesn't need to engage in selective transport. If the activated synapse is in fact marked in some way that allows it to use the newly made mRNA and protein, these can be flooded out to the entire neuron and only the marked synapse will be able to take advantage of it. This is referred to as the synaptic tagging hypothesis (Frey and Morris, 1997).

The cultured neuron experiment goes as follows: a bifurcated *Aplysia* sensory neuron extends processes in opposite directions to two separated motor neurons (Martin et al., 1997). Applying serotonin just once to either of these naïve synapses would cause only short-term facilitation. To induce long-term facilitation (the synaptic change presumed to underlie long-term memory), serotonin must be repeatedly applied at one of the sensory-motor synapses. After repeated applications, that synapse, but not the opposite one, will show new growth and increased synaptic strength that relies on new protein synthesis. If after inducing long-term facilitation at a single synapse you apply serotonin just once to the opposite synapse, that opposite synapse will also undergo long-term facilitation (Frey and Morris, 1997). This can

be explained if new mRNA and proteins, generated by the initial long-term facilitation, were distributed throughout the neuron, and the second synapse's activation by serotonin enabled their capture and use (Bailey et al., 2004; Frey and Morris, 1997).

What could the synaptic mark consist of, and how would it enable the synapse to capture and use the mRNA and protein that arrived there? The answer to these may further provide an answer to one of the most curious questions in neuroscience.

How do memories outlast molecular turnover?

A neuron, like any cell, is made up of proteins, lipids, and other molecules. Except for DNA, all these molecules are continuously broken down and remade, with the most stable components lasting no more than a few months (Crick, 1984). Nevertheless, our memories can and often do last for years or lifetimes. If there's no molecule within the neuron that is similarly stable, what neurobiological trace could possibly constitute memory? Francis Crick speculated, back in 1984, that a protein system able to recruit newly made constituents, which in turn could recruit the next new constituents, even as the older constituents are degraded, is a system that could be maintained across repeated protein turnover (Crick, 1984). Moreover, if this meta-stable structure was either influenced by the synaptic mark or was the synaptic mark itself, the proteins involved in this process would constitute the biochemical substrate of memory itself.

Two synaptic proteins in particular have been proposed for Crick's biochemical trace. One is a kinase that converts phosphorylation of itself into self-perpetuating activity. The other is a protein with prion-like properties that can promote the oligomerization of its newly transcribed monomers, forming stable aggregates. The kinase, an atypical PKC (PKMζ), is, when activated by phosphorylation, able to phosphorylate further PKMζ proteins, leading to persistent self-perpetuating activity (Sacktor, 2011). The other candidate, however, is the cytoplasmic polyadenylation element binding (CPEB) protein. Discovered first in oocytes as regulators of mRNA translation (Hake and Richter, 1994), CPEB proteins bind a canonical AAUAAA sequence in the 3' UTR of target mRNAs. In Aplysia, CPEB can also be found in neurons, and serotonin pulses applied to the neuron, as mentioned above, increase the level of CPEB (Si et al., 2003a). Moreover, reducing the availability of CPEB protein in Aplysia neurons does not prevent the initiation of synaptic facilitation, but selectively interferes with facilitation's persistence—in the absence of CPEB, synaptic facilitation begins normally but quickly disappears (Si et al., 2003a).

Memory encoded in protein conformations: a prion-like mechanism of memory

CPEB's amino acid sequence suggested a mechanism for self-sustaining behavior: 48% of the N-terminal 160 amino acids are either glutamines or asparagines, resembling canonical yeast prion proteins. The N-terminal domain was able to substitute for the yeast Sup35, and conferred stable, heritable phenotypes in the Sup35 yeast prion assay; that domain also displayed two distinct functional states in a translational reporter assay (Si et al., 2003b).

The *Drosophila* orthologue of CPEB, Orb2, is similarly required for the maintenance of flies' memory and shares its prion-like amyloidogenic behavior. There are two isoforms of Orb2 in *Drosophila*, the rare Orb2A and the more abundant Orb2B, whose sequences differ only in the N-terminal region that precedes the disordered domain. Orb2A has just 9 amino acids preceding the domain, while Orb2B has 162. To carry out Orb2's normal function, moreover, only Orb2A's N-terminal prion-domain-containing region is required, while its RNA-binding domain (which Orb2B shares, see Figure 17) is dispensable (Kruttner et al., 2012). This observation suggests that Orb2A may act as a seed, triggering the oligomerization of Orb2B, and that Orb2B's mRNA-binding domain is sufficient to regulate target mRNAs required for memory maintenance. Without Orb2A's N-terminal domain, or Orb2B's RNA-binding domain, memories can be formed but soon after disappear, mirroring the phenotype seen in flies completely lacking Orb2A. (Complete Orb2 removal causes nonviability.) A point mutation that greatly reduces Orb2A aggregation also prevents the persistence of memory, demonstrating aggregation's importance in memory maintenance (Majumdar et al., 2012). The mRNA targets that Orb2 regulates include transcripts for proteins involved in neuronal growth and synapse formation (Mastushita-Sakai et al., 2010).

Until recently, however, the function served by prion-like oligomerization was unknown—the meta-stable virtue of the oligomeric state is clear enough from the perspective of a memory trace, and oligomerization is clearly required for memory's persistence, but how does the biochemical function of Orb2 change when it oligomerizes? Work from the past few years has demonstrated that in the monomeric state, Orb2 is a translational repressor—it binds mRNA targets and brings along a deadenylation complex, shortening the mRNA's poly-A tail and leading to more rapid mRNA degradation (Khan et al., 2015). This will consequently reduce the amount of protein made from the regulated transcripts. Oligomerization of Orb2, in contrast, promotes elongation of the poly-A tail, which enhances transcript stability and increases the amount of protein made from target mRNAs. It is important to note that the oligomerization-dependent increase in translation is not merely a derepression, caused by aggregation and loss-of-function of the monomeric repressor—addition of oligomeric Orb2 to an in vitro

translation extract lacking monomeric Orb2 actually increases translation above basal levels (Khan et al., 2015).

OBJECTIVE OF THE PROPOSED STUDY

Because it is both required for the maintenance of long-term memory and displays the meta-stable behavior required of memory's biochemical trace, Orb2's aggregation has been the focus of much attention. Amyloid-like oligomerization provides theoretical benefits: a local increase in concentration, a structured platform, and a highly stable configuration. In the years since Orb2 was found to form prion-like aggregates, many new classes of functional protein aggregation have been described. While amyloids remain the most stable type of protein aggregate, non-amyloid prion-like oligomerization has been reported (Hou et al., 2011), along with a variety of signalosomes, signaling complexes, and ribonucleoprotein (RNP) granules that all exhibit features of higher-order assemblies (Wu and Fuxreiter, 2016). The domains that confer assembly behavior are, for the most part, modular—the domains can be isolated or placed in other contexts and confer similar assembly properties as in their native context. An effort initially undertaken to examine the interchangeability of these domains soon morphed into a project designed to address the requirements of Orb2's oligomerization. Did Orb2's role in memory persistence require an amyloid-like, highly ordered and stable aggregation—was that actually the conformation existing in vivo—or could it be equally well served by the more fluid interactions seen in RNP granule-associated proteins? Because no structure of any full-length prion-like proteins has been determined, querying the requirements of the in vivo system may hint at what conformations Orb2 actually assumes. Conveniently, replacing Orb2's prion domain with dozens of heterologous domains in vivo has become feasible in the age of CRISPR (see Figure 17 and Genome Editing methods). Ultimately, the goal of this work is to correlate the characteristics of the chimeric Orb2 aggregates with both the molecular function and memory phenotypes seen in animals whose endogenous Orb2 prion domain has been replaced with the heterologous aggregation-prone domains.

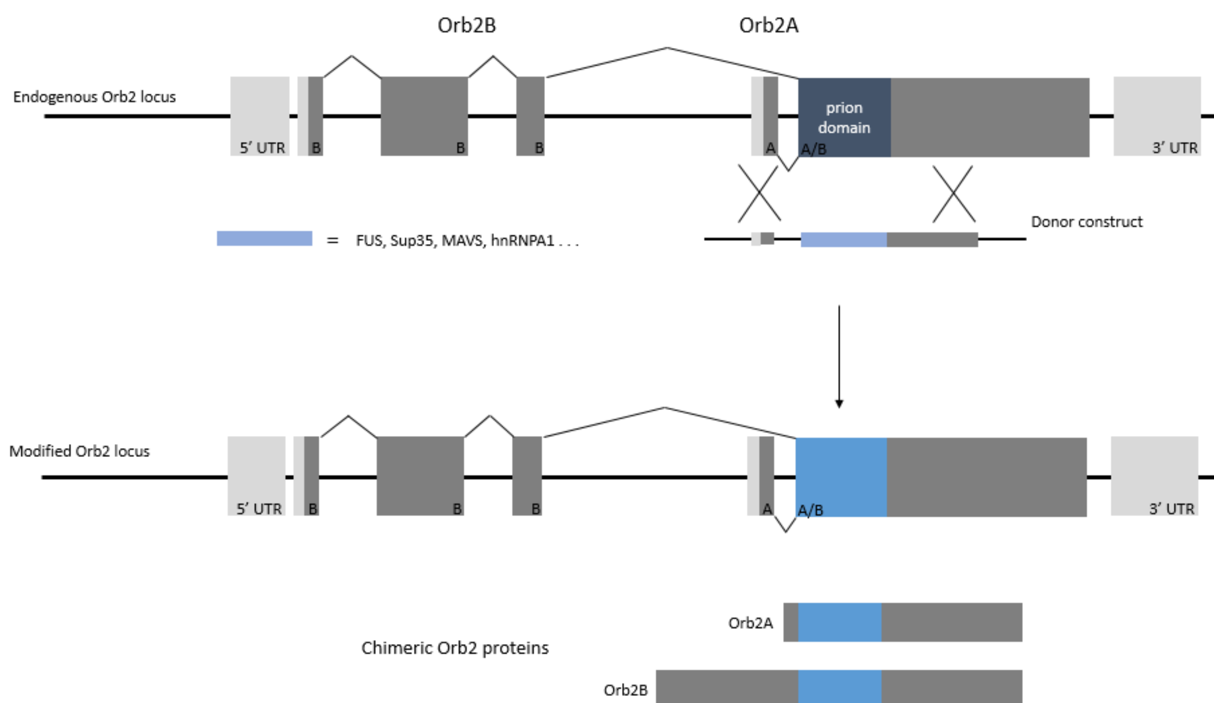


Figure 17: Orb2's genomic locus
The first three coding exons of Orb2's genomic locus are spliced together with the last (common) exon to encode Orb2B, the more abundant and less aggregation-prone isoform. Orb2A, the rarer and more aggregate-prone isoform, is generated when the small fourth coding exon is spliced together with the common exon. For this work, Orb2's prion domain was defined as the first 151 amino acids of the common exon. This region corresponds to the predicted disordered domain and was the region replaced by the various substitution domains.

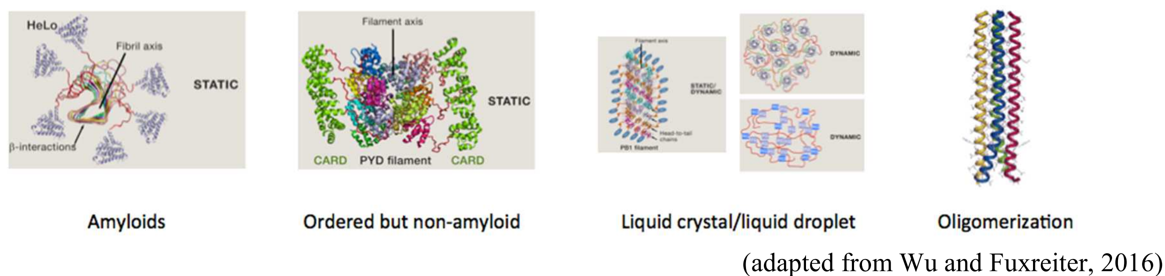
RESULTS AND DISCUSSION

SUBSTITUTION OF ORB2'S PRION-LIKE DOMAIN AT THE ENDOGENOUS LOCUS

To this end we have selected a variety of aggregation-prone domains to replace Orb2's prion-like domain (structures illustrated in Figure 18A, list of domains in Figure 18B). To insert these various domains into the endogenous Orb2 locus, we have used the CRISPR-Cas9 system to promote homology-directed repair (Doudna and Charpentier, 2014). Briefly, the CRISPR-Cas9 system is composed of a nuclease (Cas9) that can be targeted to a certain sequence in the genome by a guide RNA (gRNA), and at

that sequence Cas9 will cut both strands of DNA. In order to insert exogenous DNA sequences, or to generate precise deletions, it becomes necessary to introduce a third component to the CRISPR-Cas9 system: a donor template to guide homology-directed repair. When a cut in the genomic DNA is made in the presence of a donor template, homology-directed repair will copy the donor template sequence to fill in the cut site. If the donor happens to contain a large segment of exogenous sequence between two homology regions, that exogenous sequence can be read into (essentially inserted into) that genomic site. Our donor constructs, containing the various domains flanked by homology arms, are illustrated above in Figure 17. Flies with successfully inserted domains were screened by fluorescent red eyes, as detailed in the Genome Editing section of the Materials and Methods.

A



(adapted from Wu and Fuxreiter, 2016)

B

STRUCTURE TYPE	Domains substituted
Amyloids	Sup35, PUF4, YLR177W, SH3, beta-microglobulin, hnRNPA1, hnRNPA2, Whi3
Ordered but non-amyloid	MAVS
Liquid crystal/droplet	TIA1, FUS, TDP-43, NRP1, LSm
Oligomerization	tetrabrachion, VASP

Figure 18: The spectrum of protein aggregates
(A) Protein aggregates can take on a variety of ordered structures which vary in their stability. Amyloids are the most stable known aggregates, and contain the characteristic cross-beta sheet structure. Similarly static but lacking certain amyloid-dye binding properties, signalosomes formed by repeating six-helix bundled domains can mimic prion conversion, aggregation, and propagation. Liquid crystals, formed most commonly by RNA-binding proteins, are amorphous dynamic structures held together by a multitude of low-affinity interactions. Classic oligomerization domains, such as the tetramerization domains of tetrabrachion and VASP, serve diverse cellular functions. **(B)** A list of domains by aggregation type, to be substituted into the endogenous Orb2 locus in place of Orb2's wild type prion-like domain.

Evaluating the functional and biophysical properties of the chimeric Orb2 proteins

For these experiments to meaningfully address the structure-function relationship of Orb2, however, several issues must first be addressed. Introducing heterologous sequences into a genomic locus can influence the protein's function in a number of ways not directly connected to its conformational state. These include changes in transcription, mRNA stability, expression pattern, and the basic functions of the protein. But Orb2 is an essential gene: without at least one intact Orb2 allele, most flies die at the larval stage—just ~20% of Orb2 null mutants survive to the adult stage, and these die within a couple days. We therefore know, if homozygous flies survive and are grossly normal, that the chimeric Orb2 proteins are not completely impaired. (In at least one case, TIA1, we never recover homozygous flies. All others generated to date are grossly normal.)

Orb2 has two basic functions: binding mRNA and regulating translation. Flies showing poor memory because their chimeric Orb2 does not properly bind mRNA are not going to address the role of Orb2 oligomerization in memory. To make sure that these chimeric proteins do bind target mRNA in a manner similar to wild-type Orb2, we generated target mRNA tagged with biotin and incubated it with chimeric Orb2-containing cell lysate. (The target mRNA is a luciferase transcript whose 3' UTR is from *Tequila*, a well-characterized target of Orb2 (Khan et al., 2015; Mastushita-Sakai et al., 2010).) Subsequently pulling down the biotinylated mRNA and checking whether the chimeric Orb2 comes along with it gives an indication of whether the new Orb2 proteins still bind mRNA (Figure 19). An earlier paper mapped several nucleotides that were required for Orb2 to bind its 3' UTR target sequence—mutate these and Orb2 will no longer bind. Demonstrating that chimeric Orb2 proteins are similarly dependent on this sequence will further establish that chimeric Orb2 proteins behave similarly to wild type Orb2. Likewise, the chimeric Orb2's ability to interact with its deadenylation-complex partners (e.g. cg4612) is another basic requirement (Khan et al., 2015). Whether oligomers of chimeric Orb2 should still interact with cg4612, the interacting partner of wild type oligomers, remains unresolved: if the specific structure of Orb2 oligomers is required for the interaction, cg4612 may not be recruited, and that could explain any observed memory defect. If, on the other hand, it is simply a local concentration effect or some similar interaction between Orb2 oligomers and cg4612, the interaction may still occur. Whether lower affinity interactions between chimeric Orb2 proteins can persist, regardless of their ability to interact with the polyadenylation complex, is another unresolved question. Finally, the oligomerization of Orb2 in the activated synapse is presumed to be activity-dependent, but it remains unclear exactly how that activity changes Orb2's conformation. There is, for example, some evidence for phosphorylation increasing Orb2A's half-life, thereby increasing the concentration of Orb2A and perhaps prompting spontaneous conversion into the amyloid-like oligomer (White-Grindley et al., 2014). If in fact something like

phosphorylation is how Orb2's conformation is controlled, and we are removing and substituting the stretch of amino acids critical for this phosphorylation event, we may never find a domain that can functionally substitute for Orb2. And of course this holds for any other amino-acid-sequence-specific modification. One indication that this may not present such a hurdle, however, is data from the Keleman group. In one experiment, they replaced the Q-rich region of Orb2, 54 amino acids, with the prion-like domain of Aplysia's CPEB (Kruttner et al., 2012). This rescued courtship memory, whereas replacement with the Ure3 prion domain did not. Whatever causes the activity-prompted conformational switch in Orb2, it seems likely that it does not require the wild-type sequence of Orb2's prion-like domain.

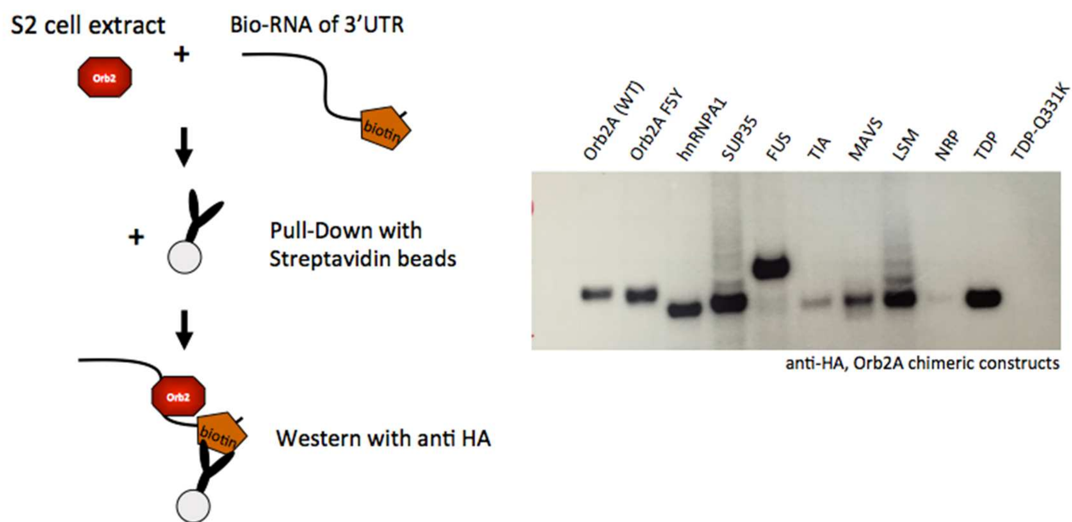


Figure 19: Details of the RNA-binding assay
Chimeric Orb2 proteins are transfected into S2 cells and given 48 hours to express. Cell lysate is then made and incubated with biotinylated mRNA constructs containing the 3' UTR sequence of an Orb2 target mRNA, Tequila. After incubation, streptavidin-coated beads are incubated with the lysate-mRNA mixture, washed, and eluted. After elution samples are run in a polyacrylamide gel and the membrane is probed for the chimeric Orb2 proteins.

Biophysical analysis

Further uncertainty arose from whether these ectopic aggregation-prone domains could exert their effect on the rest of the Orb2 protein, or whether Orb2 maintains a certain conformation regardless of the domain substituted. If the latter, all chimeric Orb2A or Orb2B aggregates should display very similar behavior. To begin addressing the aggregate characteristics of chimeric Orb2s, we have used the fluorescence recovery after photobleaching (FRAP) technique, a measure of protein exchange into and

out of aggregates (or, in the absence of aggregates, the mobility of fluorescently tagged proteins into and out of a cytosolic region). A region of interest is drawn to encompass part or all of a selected puncta, and all fluorescent molecules within that region of interest are photodestroyed by laser. Images of the darkened area are taken continuously, and any eventual fluorescence recovery in that region of interest is necessarily due to entry of new fluorescent molecules (molecules that were outside the region of interest at the time of bleaching). High recovery ('high FRAP') indicates more dynamic exchange, while a more stable aggregate would in contrast show 'low FRAP', or lower fluorescence recovery. By tagging these chimeric Orb2 proteins with YFP and expressing them in the *Drosophila* S2 and BG3 cell lines, we've found that chimeric Orb2 proteins display wide variation in the speed of recovery, suggesting marked differences in the nature of the aggregates (Figure 20).

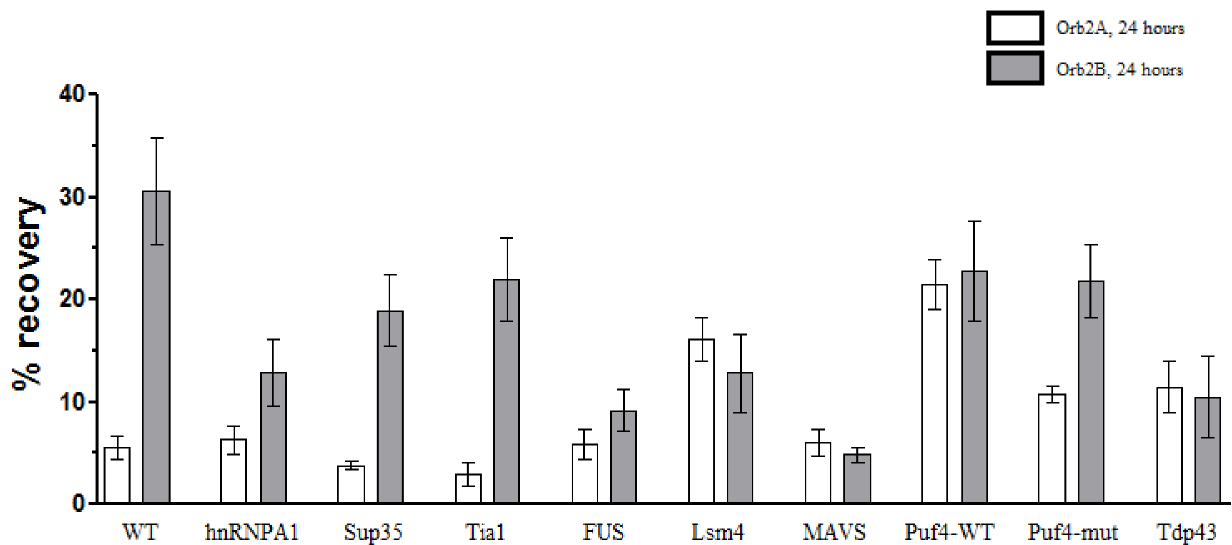


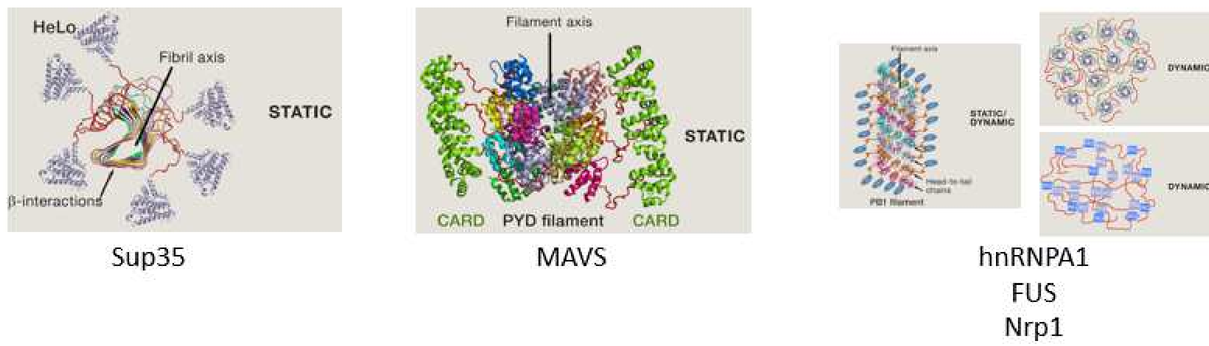
Figure 20: FRAP recovery of the chimeric Orb2 proteins
Different chimeric Orb2 proteins, 24 hours after transfection, show a range of recovery, indicating that these substituted domains are changing the aggregation properties of Orb2.

BEHAVIORAL ANALYSIS

The most stringent assay, and what we are ultimately interested in, is whether these chimeric Orb2 proteins can support long-term memory. In the handful of flies trained so far, it appears that two

domains, the FUS disordered domain and Sup35's prion domain, are able to substitute for Orb2's prion-like domain and support 48-hour memory (Figure 21). In contrast, replacement with MAVS and Nrp1 domains cause memory defects.

A.



B.

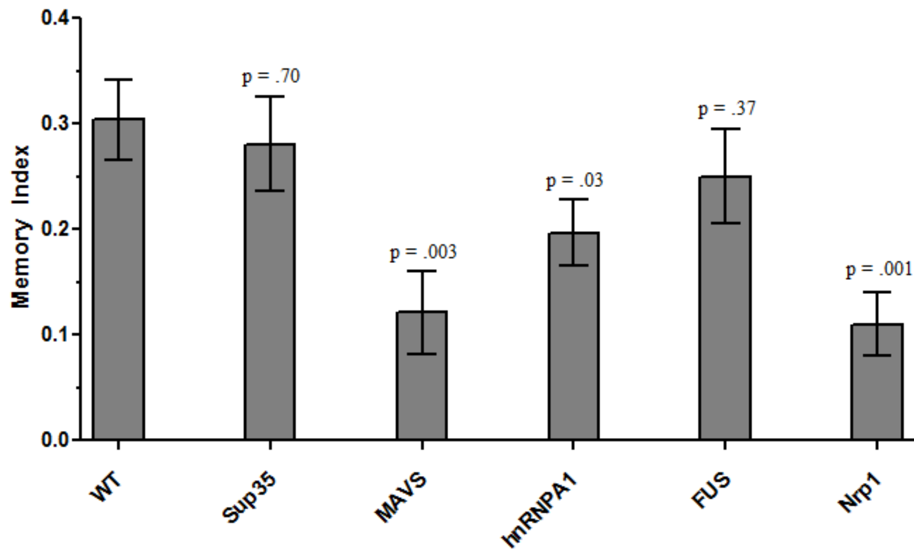


Figure 21: FUS's disordered domain and Sup35's prion-like domain can substitute Orb2's prion-like domain.

(A) Sup35 forms an amyloid structure in yeast and in vitro; MAVS forms a non-amyloid but stable structure; hnRNPA1, FUS, and Nrp1 are all members of RNA granules, and demonstrate more low-affinity and fluid-type interactions. **(B)** While FUS and Sup35 can support flies' memory at 48 hours, MAVS and Nrp1 show greatly reduced memory scores. hnRNPA1 shows an intermediate phenotype. P values are calculated by comparing the fly lines to WT using individual t-tests.

Ultimately, the goal is to identify structural characteristics that are required for Orb2's role in memory maintenance. We therefore need to see variation across several structural assays, and similar variation in long-term memory, and hope that there are obvious correlations between the two. To date, we do see variation in both FRAP and 48-hour memory, but because of the limited number of fly lines currently in hand we are not yet in a position to draw conclusions. (We will triple the number of fly lines trained in the next two months. We should similarly be able to have a much broader approach to aggregate structure, including resistance to degradation or denaturants, and potentially chemical footprinting or antibody/epitope footprinting. The search for further structural assays to characterize the conformations of these chimeric proteins is currently ongoing, and suggestions are welcome.)

APPENDIX A. MURASHKA

Orb2 and its *Aplysia* and mouse homologues share low-complexity disordered domains that underlie the persistence of memory, seemingly through a prion-like conversion process. Remarkably, unlike other known amyloidogenic and prionogenic proteins, Orb2's non-pathogenic prion-like conversion is required for normal function. The existence of this functional prion-like protein raised the question of just how broad this new biological principle could be. To determine whether there exist other proteins with similarly functional prion-like behavior, a bioinformatics analysis searched the *Drosophila* genome for common sequence characteristics of known prion-like proteins. Sequence characteristics that were used in this search include a low folding probability; low sequence complexity; sequences rich in asparagine and glutamine residues; a bias for both low charge residues and histidines, serines, or glycines; a high probability of being located at either the amino or the carboxy terminus; and the likelihood of having a mammalian homologue. The *Drosophila* genome is also enriched in opa (CAX) repeats, which primarily code for the amino acid glutamine. Therefore, in the *in silico* analysis, proteins with more than 10 continuous glutamine residues were excluded. Based on these restrictions, forty-three protein domains were identified as putatively prion-like. One of these 43 genes encoded *murashka*.

To examine whether any of these proteins possess the properties of a prion-like protein, a series of biochemical and cellular assays were performed; murashka was able to be tested in two (Alberti et al., 2010). In the biochemical assay, candidate proteins were purified from bacteria and examined by infrared spectroscopy to determine the types of molecular vibrations made by the molecule. Briefly, this technique is able to characterize purified proteins' secondary structures by monitoring the various frequencies of infrared light absorbed by the sample. The behavior of each putative prion-like protein was compared to the *Drosophila* Orb2 protein, a well-established prion-like protein in *Drosophila*, as well as control proteins with no known prion-like behavior, such as calmodulin (Figure 20). A substantial number of these prion-like proteins exhibited amyloid-type infrared-stimulated molecular vibrations, similar to the behavior of Orb2, supporting the idea that at least some of these proteins have the ability to form amyloidogenic aggregates (amyloidogenic aggregates are a common feature of prion-like proteins). Murashka exhibited strong amyloid-type vibrations that were maintained in the absence of its RING domain (a protein-protein interaction domain) (Figure 22).

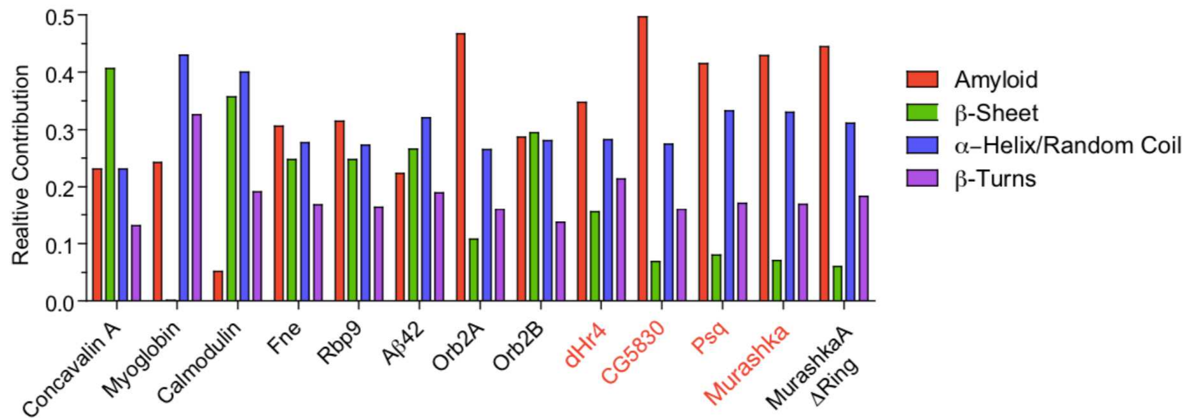


Figure 22: Murashka shows high amyloid-type vibrations in FTIR
FTIR, which shows the relative contribution of various protein folds to the overall signal, indicates that murashka exhibits high amyloid-type vibration, similar to Orb2 and several other putatively prion-like proteins

Murashka was also tested in the cell-based fluorescence recovery after photobleaching (FRAP) assay detailed in the Orb2 section. Candidate proteins were tagged with a fluorescent protein and expressed in S2 cells. A fluorescent aggregate of the candidate protein was photodestroyed by bleaching with high laser power. As described above, any increase in the aggregate's fluorescence is necessarily due to new molecules joining the aggregate, which provides a measure of the protein exchange. An aggregate of a prionlike protein, structurally very ordered, would be predicted to undergo little exchange, so low recovery is consistent with more prion-like behavior (Figure 23).

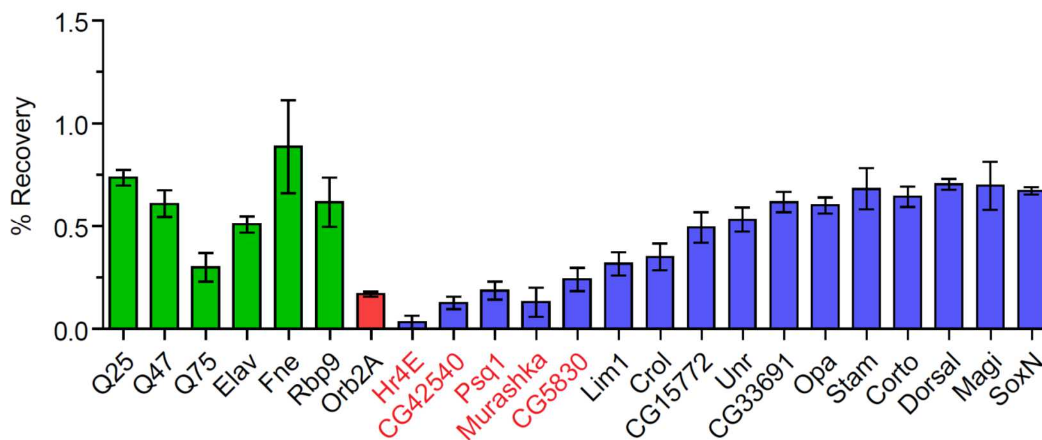
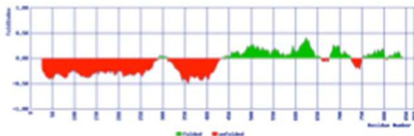


Figure 23: Candidate prion-like proteins exhibit low FRAP recovery
Five candidate prion-like proteins exhibited FRAP recovery lower than or equivalent to Orb2A, an amyloidogenic prion-like protein. Low recovery indicates less exchange with free cytosolic protein molecules, suggesting a more stable aggregate.

The results of the aforementioned assays were accumulated and analyzed. If the putative domain behaved in a prion-like manner in multiple assays, then the domain was considered prion-like. Of twenty-one proteins tested thus far, five candidates have consistently been found to exhibit prion-like behavior: cg5830, Hr4, Psq, cg42540, and murashka (Figure 23). Murashka's disordered domain, seen in red in Figure 24, is in both isoforms at the N terminus.

MuraA

MLNSNATGGNNHGGSSRRPFTRNSSSVRSQRGGGGGLGGRMVYSPHTHPHV
QQQQQQQQQQQQGQQRSSNNGNNNLWLGNSPWCNVNLNGGNSGNNNNN
GGNNNNNNGGNISSNNNSNNVNNKDANPNCHSQGQGPGNSPNLGHNHQRG
YNNGNNNRGIGNGIVNGNTNGPPDYMNYRRGVCPAQSRDYSGNGNGNNNGH
GHNGRRYMSDNLNNPSSSMNGLINSGHHQRNYNDRNLNNHFGNSEQNGGPD
HRDRAEGSSYNFMRNGGGSGGGGYGRNGSHYQHMAYGNNNGASTSTGGPG
LMGELPSGSGLSGSSLSLNNGPGGLNSDSPSRKRRRISGRPPNGPQPQHRCP
MAHMQQGSPPLRRPRLRDVATSTQQQQQQYGHQVHHPQQQQQQPPSNYHP
MHHHQPQPHYVPQQQVPPPHVQRSPWDLGSSGGSSGGATSSSSVGPILQQV
QAPPQPPSHQQAVGYPQAPPQPASLMVDLNLNQVPVNLQLRPSEPFWASFCT
YPIPAQARLAPCHLHGVYTQPFPAAPPPGLAAPPLQQQHQPLIQAQQMISQAT
LTAQQQQRDVVAIATANLGPIEAPGAHGHPHAHPHAHPHAHQLPPGIHITPLSG
AAAAAATHHLHSTAAAAAAVAAGAQITTAQQILFSSDRRTFPPHRRIPRFWTAN
HGHRHVLPPQSLAAHQAPVQIQATSGIINPGFLLNFLAMFPLSPYNQHDLSSGD
TNETENYEALLSLAERLGEAKPRGLTRNEIDQLPSYKFNPEVHNGDQSSCVVC
MCDFELRQLLRVLPCSHEFHAKCVDKWLRSNRTCPICRGNASDYFDGVDQQQ
QSQATAGAAAALSGTSGGSAGVAGTSEASAATANPQQSQAA



MuraC

MLNSNATGGNNHGGSSRRPFTRNSSSVRSQRGGGGGLGGRMVYSPHTHPHV
QQQQQQQQQQQQGQQRSSNNGNNNLWLGNSPWCNVNLNGGNSGNNNNN
GGNNNNNNGGNISSNNNSNNVNNKDANPNCDVQTSNSSGTCIYNNSKAHHHH
NNSNNNGHTQSHIHTNHSLHQGHHHPHHTQMHTPHYPHSSSPQQSNSYRQYP
PHSYSPNSPHSNGHTNSNASNNPISQRSNPQAHPNQNQNSNFYEMCTGSGGS
HTYGSMSVSLVRLNPARLCLTLGPVTPPAQRVVSFGHRSHSHSNSSSNTSSSD
QNQHNPPPRSQNHHPNPHHSQYQYQYQYQYRYQYHPPHQYHPLHSPQQQTN
RQSLPLTRTNSNSINATTFNSTNPTTNSNPTDVVNVNSCTDIIPYGSSSTSSSAS
SMLPQQQQHQINNASAPINSHSQGQGPGNSPNLGHNHQRGYNNGNNNRGIG
NGIVNGNTNGPPDYMNYRRGVCPAQSRDYSGNGNGNNNGHGHNGRRYMSD
NLNNPSSSMNGLINSGHHQRNYNDRNLNNHFGNSEQNGGPDHRDRAEGSSY
NFMRNGGGSGGGGYGRNGSHYQHMAYGNNNGASTSTGGPGLMGELPSGSG
LSGSSLSLNNGPGGLNSDSPSRKRRRISGRPPNGPQPQHRCPMAHMQQGSP
PLRRPRLRDVATSTQQQQQQYGHQVHHPQQQQQQPPSNYHPMHHHQPQPH
YVPQQQVPPPHVQRSPWDLGSSGGSSGGATSSSSVGPILQQVQAPPQPPSH
QQAVGYPQAPPQPASLMVDLNLNQVPVNLQLRPSEPFWASFCTYPIPAQARLA
PCHLHGVYTQPFPAAPPPGLAAPPLQQQHQPLIQAQQMISQATLTAQQQQRD
VVAIATANLGPIEAPGAHGHPHAHPHAHPHAHQLPPGIHITPLSGAAAAAATHH
LHSTAAAAAAVAAGAQITTAQQILFSSDRRTFPPHRRIPRFWTANHGHRHVLPP
QSLAAHQAPVQIQATSGIINPGFLLNFLAMFPLSPYNQHDLSSGDTNETENYEA
LLSLAERLGEAKPRGLTRNEIDQLPSYKFNPEVHNGDQSSCVVCMCDFELRQL
LRVLPCSHEFHAKCVDKWLRSNRTCPICRGNASDYFDGVDQQQQSQATAGAA
AALSGTSGGSAGVAGTSEASAATANPQQSQAA

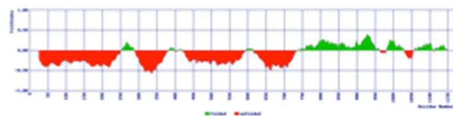


Figure 24: Murashka contains a predicted disordered domain
Murashka's protein sequence is remarkably Q/N-rich, highlighted in black. The first underlined region in muraC corresponds with the amino acids differentiating muraC (which contains them) from muraA (which does not). The second underlined region, shared with muraA, corresponds to the RING domain sequence. At bottom, the predicted disordered domain for each isoform is detailed in red, while ordered domains are in green.

UBIQUITINATION

Murashka is predicted, on the basis of its RING domain, to be an E3 ubiquitin ligase (Figures 23 and 24). Ubiquitin, a 76 amino acid protein, is a posttranslational modification known primarily for its role in targeting proteins for degradation. To mark a protein for degradation, ubiquitin is first loaded onto an E1 activating enzyme, which transfers it to an E2 ubiquitin conjugating enzyme. A further E3 ubiquitin ligase is involved, either simply to orient the E2-ubiquitin complex adjacent to the target substrate or to take the ubiquitin itself and add it directly to the substrate—in either case, ubiquitins are attached to a lysine of the target substrate. Further ubiquitination events can lead to chains of ubiquitins (polyubiquitination) or multiple lysines each with a single ubiquitin (multi-monoubiquitination), and the manner in which these ubiquitins are connected (e.g., a chain of ubiquitins connected from one ubiquitin's C-terminal to the another's Lys6, or Lys 48, or Lys63) dictates the consequence of polyubiquitination (Deshaies and Joazeiro, 2009). The number of enzymes expands at each step of the cascade, from just one known E1 enzyme in *Drosophila* to likely several dozen E2s to many hundreds of E3s (in fact, there are more predicted E3 ubiquitin ligases encoded in the human genome than kinases) (Deshaies and Joazeiro, 2009). The E3 ligases therefore provide, in combination with their partner E2s, the specificity underlying targeted degradation.

Three classes of ubiquitin ligases are differentiated by their characteristic domains and mechanisms of action: RING-domain E3 ligases facilitate the transfer of ubiquitin from the E2-ubiquitin complex to the substrate, but do not themselves bind ubiquitin as an intermediate step (Buetow and Huang, 2016). The RING domain coordinates two zinc ions to assume its distinctive cross-brace configuration, and is generally responsible for interactions with the E2 (Figure 25) (Deshaies and Joazeiro, 2009). HECT and RBR E3s, in contrast, form bonds with ubiquitin in the process of transferring it to the substrate. Murashka contains a RING domain, and on that basis is predicted to be an E3 ubiquitin ligase.

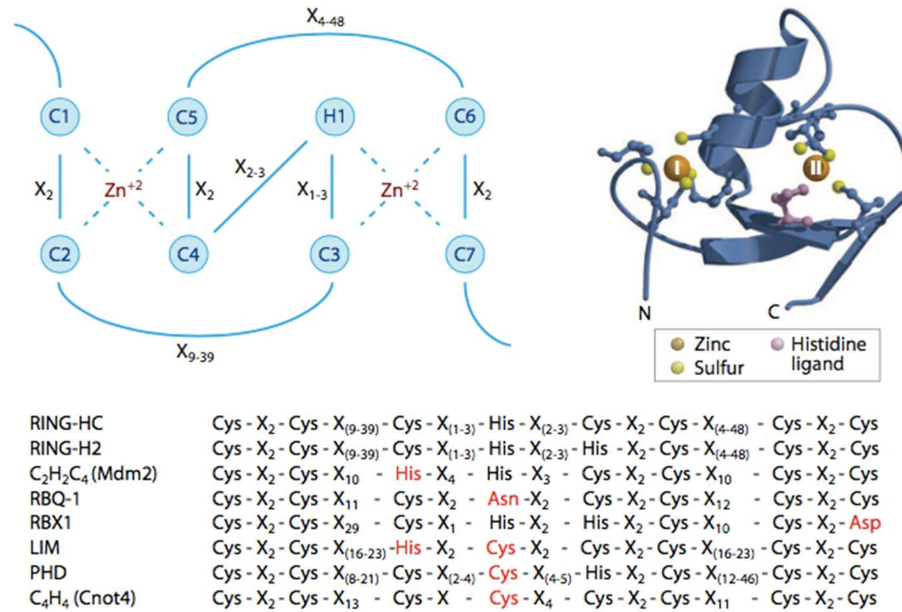


Figure 25: RING domain structure
Top panel: The RING domain in schematic form, and crystal structure (Deshaies and Joazeiro, 2009).
Bottom panel: Homology of domains similar to RING domains.

While it may seem surprising that ubiquitination, known for targeted degradation of proteins, could play a role in synaptic plasticity—if protein synthesis is so important, what's the point of degrading proteins?—though one can imagine that it isn't the absolute level of protein in the synapse that matters, but instead the relative protein composition. Degrading a protein responsible for removing AMPA receptors from the synapse, for example, would allow receptor accumulation and would thereby increase synaptic strength (Mabb et al., 2014). Pharmacological inhibition of the proteasome machinery has shown diverse memory phenotypes, including improvement of fear learning and inhibition of fear extinction (Mabb and Ehlers, 2010). Notably, both consolidation and reconsolidation have been found to require proteasome activity, and the susceptibility of recalled memories to protein synthesis inhibitors (the process of which is termed 'reconsolidation') is mitigated in the presence of proteasome inhibitor. Synapse formation and dendrite morphology are also extensively regulated by ubiquitin ligases (Mabb and Ehlers, 2010; Yamada et al., 2013). It is no surprise, then, that the ubiquitination machinery has been implicated in several neurological disorders: Angelman syndrome is due to maternally inherited loss-of-function mutations in *UBE3A*, a HECT E3 ubiquitin ligase; *UBE3A* has also been linked to autism (Mabb

and Ehlers, 2010). Mutations in parkin, a RING E3 ubiquitin ligase, are responsible for certain familial forms of Parkinson's disease.

The role played by an E3 ligase depends on its target proteins, and the consequence of those targets' ubiquitination. (While ubiquitination is canonically associated with proteasomal degradation, ubiquitination can serve many other functions, including changes in subcellular localization and protein function and assembly (Deshaies and Joazeiro, 2009).) Identifying murashka's role in long-term memory, and the likely consequences of its putative oligomerization, therefore depend on identifying its substrates and how its interactions with them are altered by oligomerization. While the likeliest scenario may be that murashka's activity is dampened or amplified in an activity-dependent manner—leading either to stabilization or degradation of its targets—it remains possible that its ubiquitination activity instead changes target localization or influences signal transduction (Deshaies and Joazeiro, 2009).

RESULTS

Murashka, named for one of Pavlov's dogs, was first identified in a paper detailing the hits of a memory screen using transposon mutagenesis (Dubnau et al., 2003). The P element transposon landed upstream of murashka's coding sequence, and mutant flies displayed normal learning and sensory behavior but had strikingly lower memory one day after training, compared to controls. Tests of the same fly line by an independent group confirmed that trained mutants showed roughly half the memory of normal controls (Akalal et al., 2011). Further, murashka was found to be one of Orb2's targets—Orb2 binds mura's 3' UTR—which is compatible with murashka's role in long-term memory (Mastushita-Sakai et al., 2010).

With the above features of murashka in mind, we have set out to identify whether murashka's predicted prion-like domain indeed confers prion-like behavior in vivo, and, if so, how that prion-like behavior influences its role in long-term memory formation.

To confirm that murashka deletion mutants indeed lack memory, as suggested by transposon mutants and RNAi studies, we completely removed the first coding exon of murashka using flanking CRISPR gRNAs (Figure 26A). Murashka mutants were trained in the courtship paradigm, a form of learning whereby males attempting to court are repeatedly rejected and subsequently suppress their

instinctive mating behavior (the evolutionary conservation of which is not the subject of this thesis). These flies show relatively normal short-term memory, suggesting intact sensory and behavior circuitry, but have dramatically reduced long-term memory (Figure 26B, left). Training in the olfactory paradigm training shows similar defects in long-term memory at both 24 and 48 hours (Figure 26B, right).

A.

B.

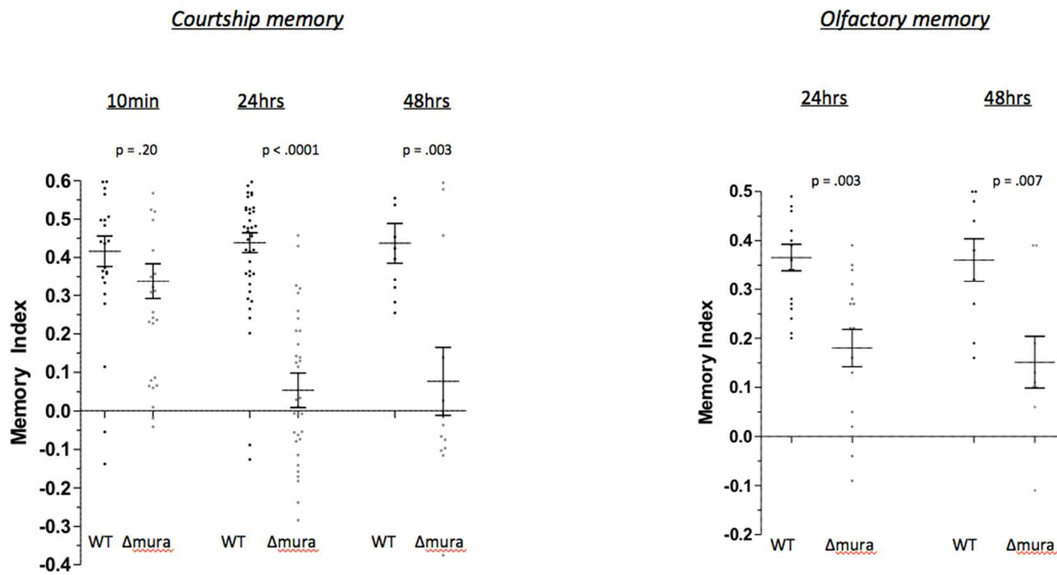
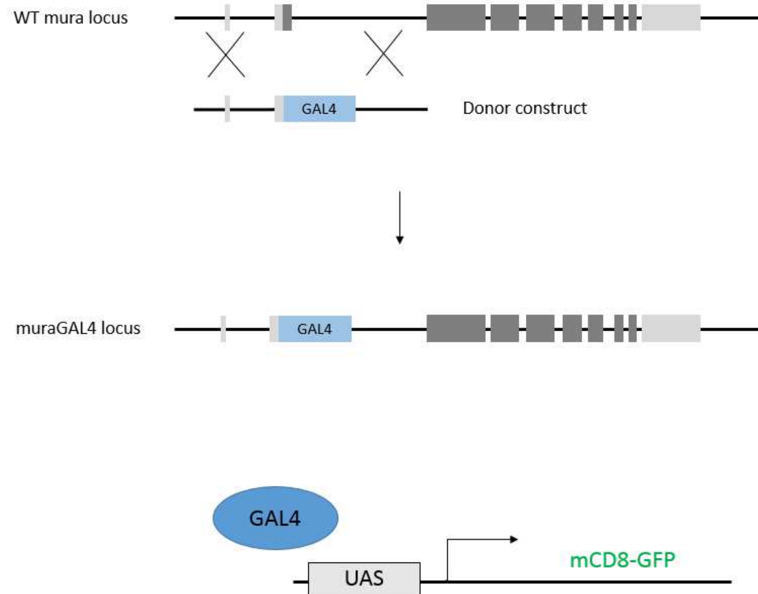


Figure 26: Murashka mutants have impaired long-term memory
(A) Left panel: Schematic of the murashka genomic locus. Mutants were generated by designing CRISPR gRNA sequences that flanked the first coding exon (red box and red highlight). Right panel: Sequence of mura's first coding exon (in grey). The sequence removed by CRISPR deletion is in lowercase letters. **(B)** Murashka mutants

show minimal defects in short-term courtship learning, but have dramatic defects in long-term memory at 24 and 48 hours after training. Olfactory memory shows similar long-term memory deficits.

Having confirmed that murashka is in fact required for long-term memory, we set out to determine where in the fly brain murashka is expressed. This was approached in two ways: first, we replaced the first coding exon of *murashka* with the GAL4 transactivator sequence (muraGAL4, Figure 27A). The GAL4 acts as a transcriptional reporter, and, when crossed with UAS-mCD8:GFP, should label all neurons in which *murashka* is being transcribed. muraGAL4/UAS-mCD8:GFP flies show widespread and surprisingly strong expression (we had until then been unable to detect murashka protein in the fly brain on Western blot) (Figure 27B).

A



B

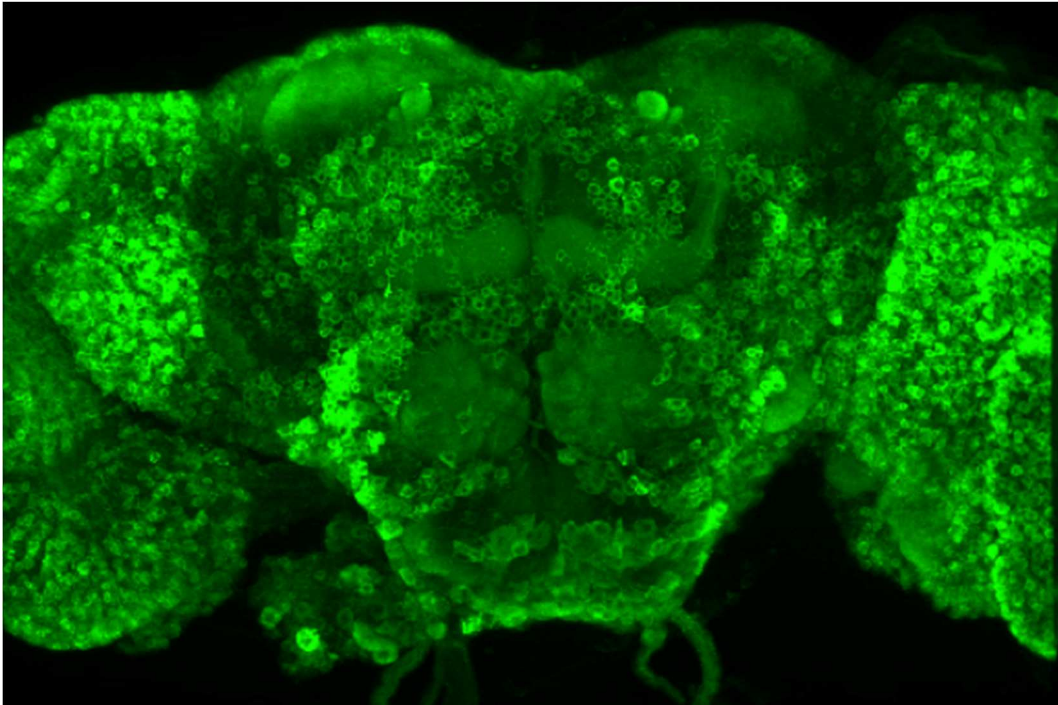
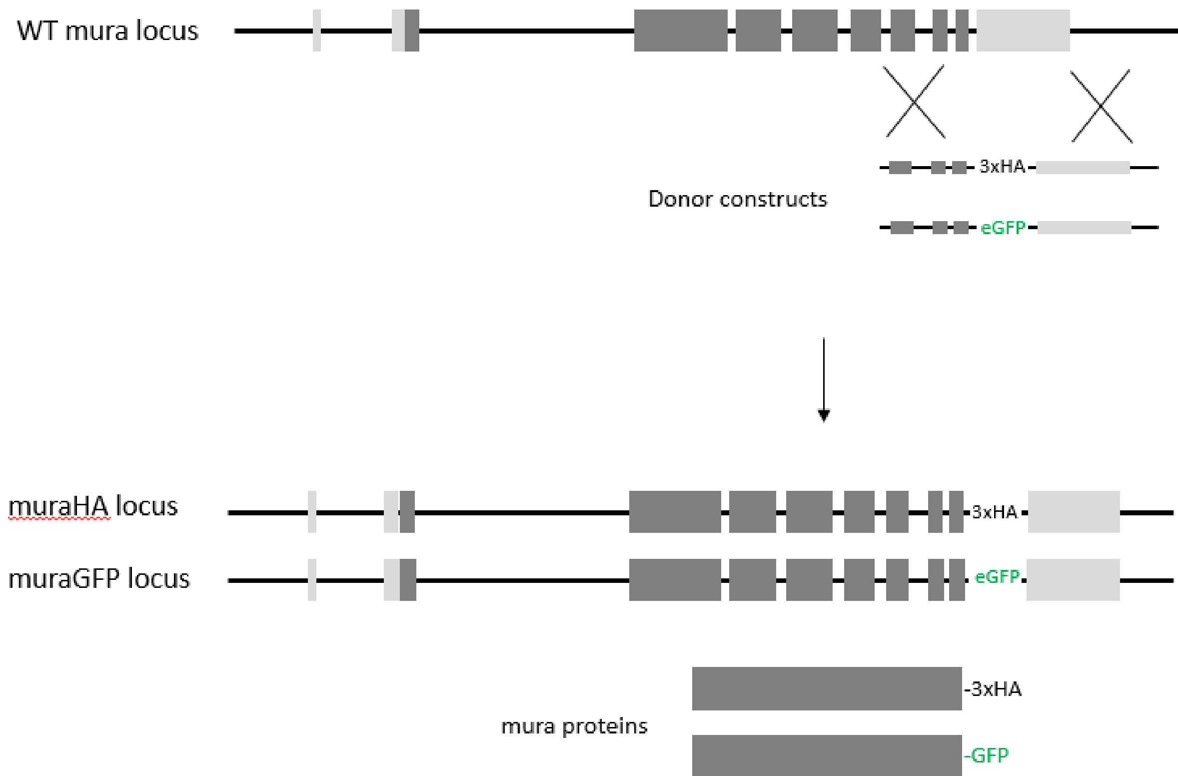


Figure 27: muraGAL4 expression pattern
(A) GAL4 knock-in at the endogenous murashka locus. **(B)** The expression pattern seen by immunostaining a muraGAL4/UAS-mCD8:GFP brain. Widespread expression of variable intensity can be seen across the brain.

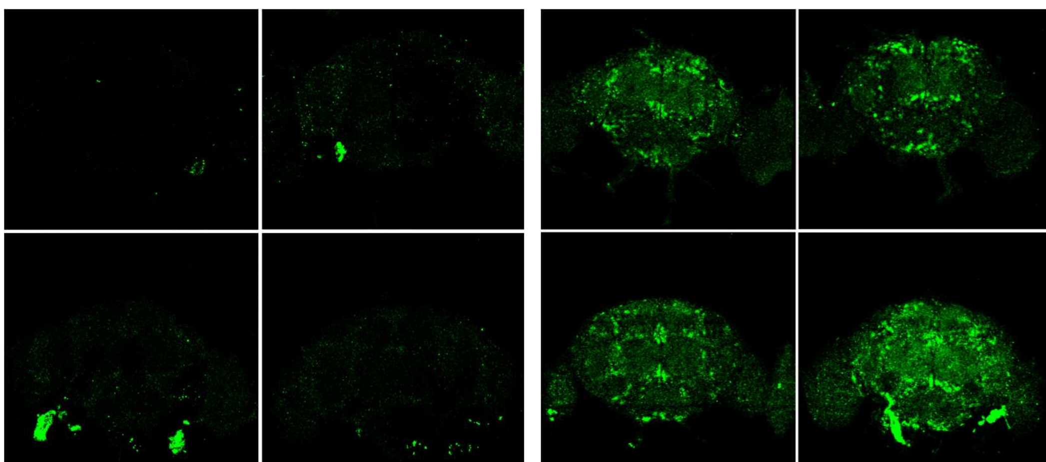
However, the GAL4/UAS-mCD8:GFP system strongly amplifies any transcriptional events, and, even supposing that we were seeing some true indication of mura's transcriptional activity, protein levels need not closely track mRNA levels (Vogel and Marcotte, 2012). To visualize the true expression of murashka, we separately tagged the C-terminus of the endogenous murashka locus with eGFP and 3xHA (muraGFP and muraHA) (Figure 28A). It quickly became apparent, imaging muraGFP brains, that murashka is expressed at very low levels, consistent with previous Western blot attempts; clear visualization of murashka expression required spectral unmixing (Figure 28B) (see Materials and Methods, Murashka). muraGFP foci appeared patchy and larger than individual cell bodies (cell bodies are on the order of 5-7 microns, while patches were typically larger than 10um). This could of course be due to expression in several adjacent cell bodies, with the space between those cell bodies unresolvable in these images. At this resolution, however, other explanations, such as extracellular localization, cannot be

ruled out. muraHA flies have also enabled detection of endogenous murashka in the fly brain (Figure 28C), though low expression still complicates several ideal experiments to identify ubiquitinated substrates.

A.



B.



C.

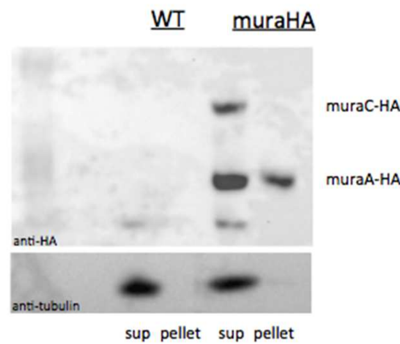


Figure 28: Generation of muraGFP and muraHA flies
(A) Murashka's genomic locus and modified loci detailing the placement of eGFP and 3xHA epitope tags. **(B)** The four left images are untagged fly brains; the four at right are of muraGFP brains. All images are sum projections taken by live-imaging and have been spectrally unmixed. (See Methods, Murashka for details.) **(C)** muraHA is able to be detected from fly head on a Western blot. Head lysate was clarified by 10min spin at 1000xG; supernatant was spun for 20min at 10,000rpm to give supernatant and pellet. The pellet was resuspended in an equal volume.

In order to determine the relevance of murashka's prion-like oligomerization, if such behavior exists, we will ideally have an in vitro assay using one or several targets as reporters. By introducing mura in its monomeric state or its oligomeric state, we could monitor the target's ubiquitination levels and perhaps identify functional differences between the monomeric and oligomeric states. Ongoing efforts to identify these substrates have so far been unsuccessful. Proteomics experiments either pulling down endogenous HA-tagged mura, or attempting to, by subtraction, determine what fly head proteins experience less ubiquitination when mura is absent, have failed to identify an actionable list of potential substrates. Neuroblast cell culture experiments, and perhaps purified murashka added into cell or tissue lysate will be attempted in the following months. Other ongoing experiments include efforts to interfere with oligomerization in vitro and in vivo to determine the behavioral consequence of mura's oligomerization. We are also attempting to purify endogenous mura to determine its biophysical properties.

Ultimately, if its disordered domain is found to confer prion-like behavior, and that prion-like behavior has an effect on long-term memory, murashka will be the second such protein identified. It would join Orb2 and suggest a broader phenomenon of proteins that stably convert their conformations in response to (presumably) neural activity. Experiments are ongoing.

MATERIALS AND METHODS

REWARD PERCEPTION AND LONG-TERM MEMORY

Fly stocks: Flies were generously shared by Dr. John Carlson (Gr43aGAL4-9/CyO; Gr61aGAL4-9/CyO; Gr5aGAL4, Gr64GAL4), Dr. Hubert Amrein (UAS-Gr43a; Gr43a^{GAL4}, with first coding exon replaced with GAL4, serving as ΔGr43a and used in crosses for behavioral training, UAS-Gr43a rescue, and ChaGAL80$^{7.4kb}$), Dr. Anupama Dahanukar (Gr61a-null mutant, Gr64a-null mutants and Gr5a-null mutants), Dr. Toshihiro Kitamoto (UAS-Shibire^{ts}), and Dr. Paul Garrity (UAS-dTrpA1). The wild type Canton-S flies were generously provided by Dr. Scott Waddell and Dr. Troy Zars. Other fly stocks were obtained from Bloomington Fly Stock Center (UAS-Kir2.1 #6595; UAS-GCaMP3 #32235; UAS-GCaMP6m #42748; UAS-ReaChR #53749; Gr64eGAL4 #57667; Gr64fGAL4 #57669; Gr64dGAL4 #57665; ΔGr64d/e #23628; ΔGr64f #27883).

Sugars: Sugars were obtained from the following sources: D-arabinose, Sigma, cat#A3131-25G, lot# SLBB3223V,100M1365V and Fisher Bioreagents, cat# BP250425, lot# 114986; L-arabinose, Sigma, cat# A3256-100G, lot# BCBB3602V,098K0164 and USB Corporation, cat# 11406, lot# 4131874; L-sorbose, Sigma, cat# 85541, lot# BCBD8834V; L-fucose, Sigma, cat# F2252, lot# SLBB1522V; L-rhamnose monohydrate, Sigma, cat# R3875, lot# BCBD8824V; D-sorbitol, Sigma, cat# S1876, lot# 017K0092; sucralose, Sigma, cat# 69293, lot# BCBF8524V; and saccharin sodium salt hydrate, Sigma, cat# S1002, lot# BCBF4560V; arabinogalactan, Food Science of Vermont, item# 026664342010.

Two-choice feeding assay using dye: The two-choice tests were performed essentially as previously described (Weiss et al., 2011): fifty 1 to 3 day old male flies were collected in groups of 50, allowed to recover for 3 days, and food-deprived for approximately 22 hours in plastic tubes (VWR) containing kimwipes wetted with 3 ml of water. (Male flies are routinely used because females' carbohydrate and protein preferences change over their reproductive cycle, and the several-day age span of flies used renders this difficult to control.) 1% agarose (Sigma) was mixed into 1M sugar solution along with red or green food dye (1%, McCormick), and 15 ul drops were pipetted into 60-well minitrays (Thermo Scientific). A hole large enough to fit a funnel was melted into the lid, and the 50 flies were allowed to feed for 5 minutes in complete darkness, with tape covering the lid hole. At the end of five minutes the color in their abdomens was assessed under a dissecting scope, and flies were counted as eating a sugar if any dye was visible in their abdomens or thorax. Flies eating a mix of the two were scored, e.g., half for L-arabinose, half for D-arabinose. Preference and detection indices were calculated as (number of flies eating sugar)/(total number of flies). To rule out the color bias in the cases of choice between two sugars, half the experiments had the colors reversed. The feeding assay was carried out for 5 minutes, instead of a period of hours, because in the context of our particular behavioral paradigm the choices made by flies over a longer time period are not relevant.

Radioactive feeding assay: Two-choice radioactivity experiments were performed as described above, with the addition of 1 uL of 1:5 diluted cytidine 5'-triphosphate [α-32P] (3000Ci/mmol 10mCi/ml ,

1MCi; PerkinElmer) into 1.5 ml 1M sugar solutions without dye; again sugars were pipetted onto the 60-well microtiter plate. After the five-minute feeding, flies were immediately placed on dry ice blocks, and five flies chosen at random were placed in each scintillation vial (Denville Scientific), homogenized, covered with 5 ml LSC-cocktail (ScintiSafe, Fisher Scientific), and counted by scintillation counter (LS6500; Beckman Coulter).

Video monitoring of feeding assay: Video monitoring of feeding flies was performed using webcams (C160; Logitech). Four colorless drops of 1% agarose and 1M sugar solution were placed on an empty 35 mM petri dish (Falcon), one in each quadrant; two were L-arabinose and two were D-arabinose. Video was recorded for 30 minutes; trials in which the flies never found the sugars were discarded from analysis. Once the fly encountered a sugar solution the behavior for next 5 minutes were quantified. We also examined the preference for other sugars, including sweet versus non-sweet sugars, to ensure that the experimental conditions did not influence the flies' choices.

Antibiotic feeding: Antibiotic experiments were carried out by placing approximately fifty 1-3 day old flies into plastic tubes with kimwipes and 3 ml of either 1M sucrose or 1M sucrose with 100 ug/ml kanamycin, 500 ug/ml ampicillin, and 50 ug/ml tetracycline, for 24 hours. The antibiotic concentrations were chosen based on previously published work (Brummel et al., 2004; Ridley et al., 2013; Sultan et al., 2001). Flies were subsequently transferred to tubes with either 3 ml water or 3 ml water with 100 ug/ml kanamycin, 500ug/ml ampicillin, and 50 ug/ml tetracycline, for another 22 hours. They were then trained with 1M L-arabinose as described below.

Survival assay: Survival curves were generated by placing fifty 3-5 day old flies in plastic tubes with kimwipes soaked in 2.5 ml of 1M sugar solution. For each sugar solution tested, ten individual tubes were tracked, thus n = 10 for each solution. The number of dead flies was counted at 12, 24, 36, 60, and 72 hours.

Appetitive-olfactory conditioning: Olfactory training was carried out largely as previously described (Krashes and Waddell, 2011): 1-3 day old flies were made hungry by placing groups of 50-70 flies in plastic tubes with kimwipes and tap water (time of starvation was determined by mortality rate: approximately 20-24 hours for homozygous lines; 24-30 hours for heterozygous crosses). 47 ul of 4-methylcyclohexanol (MCH; Sigma) and 42 ul of 3-octanol (OCT; Alfa Aesar) were separately diluted into two bubble humidifiers (B&F Medical) each containing 50 ml of mineral oil (Fisher Scientific); bubble humidifiers were connected in parallel by ¼-inch clear PVC tubing (VWR). 8 cm x 10 cm rectangles of filter paper (410, VWR) were soaked in water or 1M sugar solution, and allowed to dry until the paper was damp, then rolled to fit tightly into the training tubes. Groups of 50-70 flies were moved into the t-maze, then into the water tube for 2 minutes while MCH odor passed through, moved back to the holding chamber in the t-maze for 30 seconds, then moved to the sugar tube for 2 minutes while OCT odor was flowing through. The next group of flies was trained reciprocally, where OCT was paired with water and MCH with sugar. Unless otherwise specified, after training flies were fed for 4 hours and restarved until testing 24 hours after training. Flies were tested by being given a choice between OCT and MCH in tubes with no filter paper; test duration was 2 minutes. Short-term memory was assayed two

minutes after training. Memory index = [(number of flies in reward odor – number of flies in unrewarded odor)/(total number of flies)]. A memory index was calculated for each of the two reciprocal trials and then averaged; this average constituted an *n* of 1. Sucrose was frequently used as a daily standard, thus the large numbers of sucrose trials. For experiments with two or more controls, the experimental line was trained in parallel with one of the controls, and then again trained in parallel with the other control—thus the large *n* for both ChaGAL80 and ReaChR experiments.

Optogenetic and TrpA1 stimulation: Earlier attempts at similar light-activated neuronal firing had used receptors whose ligands were blocked by photoremovable inhibitors; pulses of light removed the inhibitor, activated the receptor, and caused neurons to fire (Lima and Miesenbock, 2005).) Genetic targeting of channelrhodopsin, whether by virus or germline modification, allowed a defined (i.e. only neurons expressing gene X will express ChR2) subset of neurons to be turned on and off at will merely by shining light on them. Previous methods for activating neurons had often relied on electrical stimulation, which activated all neurons in the vicinity of stimulation. Two other methods involve receptors designed to recognize an exogenous compound not normally found in the animal (Urban and Roth, 2015), and heat-activated Trp receptors that when ectopically expressed cause neurons to fire in response to higher temperatures (Hamada et al., 2008; Pulver et al., 2009). Neither of these two methods allows the exact, second- to sub-second control that optogenetics and electrical stimulation permit, but both have nevertheless been useful in a wide variety of settings.
Optogenetic activation was performed using the same hardware as previously published (Inagaki et al., 2014), except that two rows of six LEDs each were aligned parallel to the tube, 2 cm away, at 90º angles to each other. To minimize behavioral artifacts caused by strong visual stimulation, the red (627 nm, 161 lm @ 700 mA) Rebel LEDs were chosen, and the stimulation protocol (pulse width, intervals, and duration) was controlled by Arduino board and Arduino computer language. For dTrpA1 experiments, the relevant training tube was preheated to 31 ºC, and during training was wrapped in a ReptiTherm Under Tank Heater (RH-4; Zoo Med Laboratories); the temperature was held constant (at 31 ºC) by an electric temperature control with probe placed in between the wrapped layers (A419; Johnson Controls). The heater temperature required to maintain an internal tube temperature of 30 ºC was determined empirically.

Statistical analysis and number of trials (*n*). All statistical analysis was performed using Graphpad Prism 5. All of the data met the assumption of homogeneity of variance, therefore unpaired two-tailed t-test or one-way analysis of variance (ANOVA) was performed, with Tukey post-hoc test between pairs of samples. ANOVA tests for significance were performed a probability value of 0.05 and more stringent values are listed in each figure where applicable. For all experiments, each *n* is considered a biological replicate; separate trials used independent samples of genetically identical flies. For two-choice experiments, a single *n* constitutes a population measure generated from 50 male flies. The preference index indicates the proportion of flies eating the sugar, which was determined by scoring visible color in the abdomen or thorax. For video monitoring, each *n* constitutes a single fly. For survival curves, each *n* is a population measure generated by 50 flies placed in a tube with 1M sugar. Percent survival indicates the percentage of flies alive at each timepoint. For olfactory training experiments with sugar, heat, and light: one trial consists of giving a group of approximately 50-70 flies water and 3-octanol for two minutes, waiting 30 seconds, then giving sugar and 4-methylcyclohexanol for two minutes. Another

group is trained with water and 4-methylcyclohexanol, then sugar and 3-octanol. Memory indices are calculated for each of these two trials and averaged. This average constitutes a single *n*, which is approximately 100-140 flies. Based on previous and ongoing experimental effect sizes, 8-10 of these double trials were generally judged to be adequate for memory experiments, unless effect sizes were strikingly large or variable. The more dramatic effect sizes and smaller variability of preference assays allowed a smaller number of trials, generally 4. In all long-term memory experiments, experimental manipulations for which a negative result was plausible or expected were always trained alongside a positive control. This is the reason for conspicuously large numbers of trials with sucrose and L-arabinose compared to other sugars or manipulations. Similarly, for experimental groups needing to be compared to two or more controls, the experimental group was first trained alongside one of the control groups, and then again trained alongside the other control group (s). This is the reason for large numbers of trials in, for example, the ChaGAL80 and ReaChR experiments.

Immunostaining: Tissues were dissected in PBS (Sigma), and fixed in 4% paraformaldehyde (Electron Microscopy Sciences) in PBS-Triton .3% (PBST) (Sigma) for 1-2 hours. They were washed in PBS-Triton .3% 5 times for 15 minutes each time, and blocked in PBST with 10% normal goat serum (Vector Laboratories) for two hours. Rabbit anti-GFP IgG (MBL International Corporation) was diluted 1:1000 in the blocking solution and centrifuged at 14,000 r/min for 10 minutes at 4 ºC. Tissues were incubated with primary antibody overnight at 4 ºC, then washed again with PBST for 15 minutes, five times. Anti-rabbit IgG Alexa Fluor 488 (Life Technologies) was diluted 1:1000 in blocking solution, and incubated with the tissues overnight. Tissues were again washed five times, and mounted in Vectashield (Vector Laboratories) on slides with doubled clear reinforcement labels (Avery); No. 1 ½ coverslips were used (VWR). Images were acquired on a Zeiss Pascal confocal microscope with a Plan Apochromat 20x 0.8NA objective. GFP fluorescence was excited at 488nm and emission was collected through a 505-530 nm bandpass filter.

Calcium Imaging: Tissues from Gr43a^{GAL4} x UAS-GCaMP3 or UAS-GCaMP6med flies were prepared largely as described previously (Miyamoto et al., 2012). Two to seven day old flies were used. All tissues were dissected in Ringers solution (5 mM HEPES, 130 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$); legs were removed from the fly, placed on a 50 mm glass-bottomed dish No 1.5 (Mattek), and immobilized with a 1.5 ul drop of 2-hydroxyethylagarose (Sigma). After the agarose firmed, 20 ul of Ringers was added to cover the leg. In D- versus L-arabinose comparisons, both front legs of the fly were used as matched controls. Brains adhered to the dish without need for agarose when placed into a 30 uL bubble of Ringers. Proboscis imaging was performed with the proboscis upside down on the plate, so that the dorsal proboscis was contacting the dish; the proboscis was immobilized with 1.5 ul of agarose and again covered with 30 ul of Ringers solution. Only one sugar was tested per tissue sample. Images were collected at least 40 seconds before sugar was added; sugar was added at 2x concentration, in the same volume as the Ringers covering the tissues. Because the training paradigm uses high concentrations (up to 1M) of sugar, we used 500 mM sugar concentrations for the leg and proboscis imaging. However, for the brain, 500 mM appeared to cause osmolarity-induced shrinking, so brain imaging used 100 mM sugars. Leg imaging was performed at approximately one stack per 5-7 seconds; proboscis imaging at approximately one stack per 13 seconds; brain imaging at approximately one per 14 seconds. Only tissues

that showed a response were used in analysis, though tissues that didn't respond were checked for viability by adding fructose as a positive control. Images were acquired on a Zeiss Pascal confocal microscope with a Plan Apochromat 20x 0.8NA objective. GFP fluorescence was excited at 488 nm and emission was collected through a 505-530 nm bandpass filter. For calcium imaging of the leg with $\Delta Gr43a^{GAL4}$ x UAS-GCaMP3 and $\Delta Gr43a^{GAL4}$; UAS-GCaMP3 flies, images were acquired on an Ultraview Vox (PerkinElmer) with a Plan Apochromat 20x 0.8NA objective at approximately one stack per 10 seconds; GFP fluorescence was excited at 488 nm and collected through a 525-550 nm bandpass filter. Analysis was performed in ImageJ (NIH) using custom in-house plugins: z-stack images were sum-projected and camera background was subtracted by selecting a region of interest away from the tissue. Where needed, the StackReg registration plugin was used to minimize movement artifacts (Thevenaz et al., 1998). Measurements were always taken by encircling cell bodies. In tissues with more than one neuron visible, the response of each neuron was analyzed separately and then averaged to generate an average response for that single tissue; this average constituted a single *n*, and was used with others to generate average response curves and peak $\Delta F/F_o$. Peak $\Delta F/F_o$ measurements were made by taking the first peak value, and dividing by the average of five timepoints immediately preceding the rise. To generate normalized fluorescence curves, individual tissue averages were aligned by the first timepoint of the rise. Curves for leg and proboscis were linearly resampled at 3 seconds; brain at 5 seconds. Curves were then min/max normalized, and average trajectories were calculated. Error bars were calculated as standard error in the mean. Average curves were plotted in GraphPad Prism 5.

CAFÉ assay: One day old Canton S adult flies (males and females) were transferred to fresh standard food medium for one day and then starved (with free access to water) for 18-22 hours. These flies were then transferred by groups of 20 into plastic boxes (Sellier et al., 2011). Each box had a row of 5 capillary tubes (5 µl minicaps, Hirschmann LaborGeräte, Germany), filled with a dilution of sugar mixed with a red dye (erythrosine 0.374 mg/ml; Sigma France). The concentrations of sugar (L- and D-arabinose, Sigma, France) were: 1 M, 100 mM, 10 mM, 1 mM and 0 mM. Each box was monitored with a webcam (HD Pro C920 or QuickCam Pro 9000, Logitech). The boxes and cameras were housed in a climatic chamber maintaining a temperature of 25°C and 80% H.R. (DR-36 VL, CLF Plant Climatics GmbH, Germany). For each box images were acquired at a rate of one image/ min for 2 h using the software VisionGS, Germany. The stack of images was then transferred to ImageJ (Abramoff et al., 2004) and the liquid level of each capillary was analyzed using a custom Java plugin, and subsequently transferred to Excel. Results are expressed as the mean of the change of the liquid level in each capillary (D-arabinose: n = 12; L-arabinose: n=10 boxes). Error bars are computed as the standard error to the mean (s.e.m.).

Electrophysiological Recordings: Tip-recording was performed as previously described (French et al., 2015). Briefly, adult flies (3-4 d old) were anesthetized on ice and immobilized on a putty platform (UHU stick), using thin stripes of tape. They were then disposed under a stereomicroscope (MZ12, Leica) and specific sensilla from the proboscis or from the legs were stimulated and recorded, using a TasteProbe amplifier (DT-02, Syntech, Germany; Marion-Poll and van der Pers, 1996) connected to a general purpose amplifier (CyberAmp 320, Data Translation, USA) which further amplified (x100) and filtered the signal (10Hz-2800 Hz). The stimulus electrode contained tricholine citrate (TCC 30 mM), in order to allow an electric contact to be established with the sensillum and to inhibit firing activity arising from

water-sensitive cells (Wieczorek and Wolff, 1989). A reference electrode was connected to the abdomen of the fly, using a drop of electrocardiogram gel. Each stimulation lasted 2 s and was digitized at 10 kHz, 16 bits during 2 s (DT9818, Data Translation, USA). The data acquisition, spike detection and sorting was performed under a custom program, dbWave. The results were subsequently transferred to Excel, and expressed as the mean (n=8-15 measures). Error bars were computed as the s.e.m.

ORB2'S CONFORMATIONAL SWITCH

Vectors: To generate Orb2A and Orb2B expression vectors for transfection into S2 *Drosophila* cells, TOPO-D vectors containing the Orb2 CDS were used as templates for PCR that used primers beginning immediately after the prion domain and ending just prior to it (Orb2A topo F: ACTGGCATGGTCTCAGCC; Orb2A topo R: AATGAAATTAACAAATTTGTTGTACAT; Orb2B topo F: ACTGGCATGGTCTCAGCCGG; Orb2B topo R: ACTTTTTCGATTGAAGAAACCACCGG). The domains to be inserted were chosen based on previous work that had defined or worked with the various domains (hnRNPA1, hnRNPA2, FUS, TIA1, TDP43, SUP35 from (Kato et al., 2012); Puf4 and YLR177W from ((Paul et al., 2015); LSM4 from (Decker et al., 2007)). These domains were PCR'd or synthesized using primers with Gibson overlaps
(domain into Orb2A F: ATGTACAACAAATTTGTTAATTTCATT_____________;
domain into Orb2A R: GTTGCTCCGGCTGAGACCATGCCAGT______________;
Orb2B overlap F: CCGGTGGTTTCTTCAATCGAAAAAGT____________;
Orb2B overlap R: GTTGCTCCGGCTGAGACCATGCCAGT___________) to insert them into the Orb2A/B TOPO-D vectors. Gateway LR clonase II reactions shuttled the chimeric Orb2 gene into two expression vectors, pAHW (DGRC #1095) for the C-terminal 3xHA tag used for the RNA-binding assay and pAWV (DGRC #1088) for the C-terminal Venus tag used in the fluorescence recovery after photobleaching (FRAP) experiments.

Fly stocks: The pHD-DsRedScarless vector used to insert the substitution domains into the endogenous Orb2 genomic locus is found in more detail in the Genome Editing section. The first step inserted the 3' arm into the SapI site by Gibson cloning, with F primer
CAATTTTACGCAGACTATCTTTCTAGGGTTAAACTGGCATGGTCTCAGCCG
and R primer GAGACGTATATGGTCTTCTTTTCCCGGTGCTTTCGTTTCTCGCCTCTAAAAC. The 5' arm extended up to the beginning of Orb2's prion-like domain (the first full codon of the common exon). Inserted immediately adjacent to and downstream of this 5' arm will be the various prion domains, but to construct a vector that could be used to insert any sequence at that locus, we amplified the sequence immediately following the prion domain extending to but not including the next TTAA sequence found in the Orb2 gene. In the F primer used to amplify this 'TTAA fragment' we included the AarI site and surrounding nucleotides, so that we would be able to digest the vector with AarI and easily insert any small substitution domain.

The donor pHD vector was injected, along with a pU6.2-BbsI-chiRNA vector (gRNA target sequence TCACATAGCCCCTCGTCGCC), into a *nos*-Cas9::attp40 fly. The progeny of the injected larvae were

examined for red fluorescent eyes (or red segments in the larval tail), and these individual flies were isolated and crossed with balancer flies to set up potential stocks. The founders were PCR sequenced to verify proper insertion, and then crossed with a piggybac-transposase-expressing fly (Bloomington Fly Stock #32072). The heterozygous flies were crossed with a third chromosome balancer. Because the piggyBac chromosome is marked with eye-expressing CFP, the progeny of that heterozygous/balancer cross were screened for a lack of blue eyes, and then a lack of red eyes. Flies without CFP (meaning we weren't looking at a balanced piggyBac chromosome) and without dsRed (meaning the pHD-DsRed vector had been removed) were found at a frequency, usually, of about 10-20%. These non-CFP, non-dsRed flies were single-fly crossed again with balancer flies to set up individual lines. The lines were again PCR sequenced to verify proper dsRed excision. In every case, correct excision was found within the first five lines.

RNA-binding assay: pAHW vectors containing the chimeric Orb2 CDS were transfected into S2 *Drosophila* cells and allowed to express for 48 hours. The cells were lysed (contents of lysis buffer:________________________), and the lysate was checked by Western blot for protein expression. Lysates containing the chimeric Orb2 proteins were incubated for four hours with biotinylated Tequila reporter mRNA (HOW WAS RNA biotinylated, what's the sequence of the Teq reporter (Teq 5' UTR – luciferase CDS – Tequila 3' UTR). Streptavidin beads were then added to this lysate-biotinylated mRNA mix to pull down the biotinylated mRNA. After removing the lysate, the streptavidin beads were boiled with loading dye and loaded into an SDS-polyacrylamide gel. The membrane was probed with anti-HA HRP (Sigma, H6533).

Fluorescence Recovery After Photobleaching: pAWV vectors containing the chimeric Orb2 constructs were again transfected into S2 cells. 24 hours after transfection cells expressing the YFP-tagged construct were identified by fluorescence. Depending on the construct, cells either displayed many discrete puncta, a small number of larger amorphous aggregates, or a single large aggregate. A region-of-interest was drawn to include part of the aggregates, or, if too small, a region containing several small puncta adjacent to a region with several other puncta. FRAP measurements were performed largely using the method detailed in Majumdar et al., 2012. Microscope images were acquired using a Carl Zeiss LSM-510 confocal microscope (Jena, Germany) with either a C-Apochromat 1.2 NA water objective or a Plan-Neofluar 1.3 NA oil immersion objective with a pinhole size of approximately 2 airy units and a pixel dwell time of approximately 6.4 microseconds. The red and green signals, excited with 488 and 561 nm lasers, were separated with an NFT 565 beam splitter, and collected by photomultiplier tubes (PMTs) through 505-550 and LP 575 filters. All images were taken with one-second time resolution. Five images were taken before photobleaching of a small region containing a punctate structure. Thereafter, cells were observed for 50 seconds to assess recovery of both the diffuse pool and individual puncta. Great care was taken to avoid cells in which large cytosolic rearrangements occurred during the recovery as this would invalidate any measurements of percentage recovery of puncta. The averaged recovery curves within the bleached region were fit to a single rising exponential as follows: $I(t) = A(1-\exp[-t/\tau]) + B$, where τ is the recovery time constant (inverse rate). A is the recovery amplitude, and B is the background intensity. Note that the half time of recovery is equal to 0.69τ. The tau value was fixed to 50 microseconds in each case. The more informative number for our purpose is the fraction mobile, calculated as

follows: $f_{mobile}=A/(I_{prefrap}-B)$. Here $I_{PREFRAP}$ is the average intensity in the bleach region before bleaching. We typically see around 20% of the recovery intensity coming from dynamics of the diffuse pool. Therefore mobile fractions significantly greater than this are indicative of puncta recovery, i.e., bleached puncta proteins exchanging with the diffuse pool. In these cases, the puncta clearly recovers.

Memory Training

Appetitive-olfactory conditioning: Olfactory training was carried out largely as previously described (Krashes and Waddell, 2011): 1-3 day old flies were made hungry by placing groups of 50-70 flies in plastic tubes with kimwipes and tap water (time of starvation was determined by mortality rate: approximately 20-24 hours for homozygous lines; 24-30 hours for heterozygous crosses). 47 ul of 4-methylcyclohexanol (MCH; Sigma) and 42 ul of 3-octanol (OCT; Alfa Aesar) were separately diluted into two bubble humidifiers (B&F Medical) each containing 50 ml of mineral oil (Fisher Scientific); bubble humidifiers were connected in parallel by ¼-inch clear PVC tubing (VWR). 8 cm x 10 cm rectangles of filter paper (410, VWR) were soaked in water or 1M sugar solution, and allowed to dry until the paper was damp, then rolled to fit tightly into the training tubes. Groups of 50-70 flies were moved into the t-maze, then into the water tube for 2 minutes while MCH odor passed through, moved back to the holding chamber in the t-maze for 30 seconds, then moved to the sugar tube for 2 minutes while OCT odor was flowing through. The next group of flies was trained reciprocally, where OCT was paired with water and MCH with sugar. Unless otherwise specified, after training flies were fed for 4 hours and restarved until testing 24 hours after training. Flies were tested by being given a choice between OCT and MCH in tubes with no filter paper; test duration was 2 minutes. Short-term memory was assayed two minutes after training. Memory index = [(number of flies in reward odor – number of flies in unrewarded odor)/(total number of flies)]. A memory index was calculated for each of the two reciprocal trials and then averaged; this average constituted an *n* of 1.

Male courtship suppression assay: The male courtship conditioning assay was modified from that described previously (Majumdar et al., 2012). Each male virgin was isolated right after eclosion in standard food vials. When they mature to 4~5 days old, each virgin male was paired with a freshly mated female for one to three sessions of 2 h each, with a 30 min rest period in between. During training sessions flies were kept in 16 x 100 mm culture tubes (VWR) provided with standard corn syrup fly food. Memory performance was tested with a fresh-mated female at the indicated time point in a 1 cm diameter wheel. A courtship Index (CI) was measured as the fraction of time the tested male spent chasing the female in a 10 min interval using an automated ImageJ based program. The Memory Index or courtship suppression index was calculated as: , where CI_{Naive} and $CI_{Trained}$ are the mean courtship indices for independent samples of naive and trained males, respectively.

MURASHKA

Flies: Flies with C-terminally HA or GFP tagged murashka were generated by the methods described in the Genome Editing appendix. 5' arm, CGCAGCAGAT . . . AGCCAAGCGGCC; linker and HA (or substitute GFP sequence),

ACTAGTTACCCATACGATGTTCCAGATTACGCTTACCCATACGATGTTCCAGATTACGCTTACCCATACGATGTTCCAGATTACGCTTAG; 3' arm, ACCGGAA . . . AATATACTA. CRISPR target sequence, CAGAGCCAAGCGGCCTAGAC.

Δmura flies were created using two gRNAs, AACACTTAACTATGCATCC and GAAAGGATCTCGAGTGAGT. Deletion removed all nucleotides between AAACGGAAC . . . TCCTTTCCT, removing the entire first coding exon.

Imaging

Confocal Imaging: Images of *Drosophila* brains were acquired in photon counting lambda mode on a Zeiss LSM-780 using the GaAsP detectors. Z stack tiles were scanned with 20% overlap. The tiles were then stitched using the grid/collection stitching plugin in Fiji. Control GFP and autofluorescence spectra were acquired as images and used for unmixing. Fluorescence unmixing was accomplished with in-house written custom plugins in Fiji.

Electron Microscopy: For immuno-EM, the fly brains were dissected and fixed in Cryofix (4% PFA, 0.01% Glut in PBS, pH 7.4) for at least 2 hours. The samples were then washed 3 times in PBS for 10 min each. After dehydration with gradient Ethanol and infiltrated with LR white resin, the samples were embedded and polymerized at 60C for 24 hours. Light was avoided during all above processing. Thin sections of 80nm were then cut with a Leica Ultramicrotome and arranged onto the sliver grids. The grids were stained with anti-HA antibody followed by 12 nm colloidal gold secondary antibody (Jackson Immunoresearch Labs; 1:50 dilution).

Memory training

(See methods of Orb2 prion domain substitution.)

GENOME EDITING IN DROSOPHILA

Thanks to the discovery and development of the CRISPR-Cas9 system (Doudna and Charpentier, 2014), it has become feasible to insert, remove, or modify nearly any sequence in the *Drosophila* genome (Gratz et al., 2015). The lab has so far seamlessly inserted epitope and fluorescent tags on either end of various genes, introduced specific protease cleavage sites into endogenous proteins to allow controlled degradation, and replaced the endogenous prion-like domain of Orb2 with twenty different disordered domains. We are heavily indebted to the flyCRISPR group at the University of Wisconsin-Madison for reagents and tools used in this work.

Briefly, the CRISPR-Cas9 system is composed of a nuclease (Cas9) that can be targeted to a certain sequence in the genome by a guide RNA (gRNA), and at that sequence Cas9 will cut both strands of DNA. Simple but relatively uncontrolled deletions can be made by designing one or two guide RNAs at or surrounding the region to be removed and relying on non-homologous end joining to (attempt to) repair the cut (Gratz et al., 2014). This error-prone repair often generates insertions or deletions that can cause frameshift mutations, and at some frequency will fully excise the region between two cuts. Removing, for example, the first coding exon of a gene will (ideally) eliminate expression of the protein.

In order to insert exogenous DNA sequences, or to generate precise deletions, it becomes necessary to introduce a third component to the CRISPR-Cas9 system: a donor template to guide homology-directed repair. When a cut in the genomic DNA is made in the presence of a donor template, homology-directed repair will use the donor template sequence to fill in the cut site. If the donor happens to contain a large segment of exogenous sequence between two homology regions, that exogenous sequence can be read (essentially inserted) into that genomic site.

One method to identify flies carrying this new exogenous sequence is PCR screening—either relying on a primer whose complementary sequence is found only in the exogenous insertion, or simply by an increase or decrease in the size of the amplified fragment. Another less laborious method useful in *Drosophila*, however, relies on extra exogenous sequence: by adding a small promoter and gene for a red fluorescent protein (dsRed), flies who carry the insertion of the desired sequence will also have eyes that fluoresce red. This piggyBac method (pBac) then relies on a transposase to remove the pBac vector sequence, leaving behind only the desired exogenous sequence.

The steps below illustrate two uses of the CRISPR-pBac system, one to tag the N-terminus of a protein and the other to insert a GAL4 transactivator sequence into the endogenous first coding exon of a gene.

Designing donor templates

The basic piggyBac vector we have used as a donor for homology-directed repair is pHD-ScarlessDsRed:

AGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGG
CTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGCCCACCACTTCAAGA
ACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATA
AGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACG
GGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGT
CGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGG
TTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAA
CGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCG
TTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGA
ACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGGCGAGAGTAGGGAACTGCCAGGCATC
AAACTAAGCAGAAGGCCCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTCTGTGTTGTAAAAC
GACGGCCAGTCTTAAGCTCGGGCCCCCTGGGCGGTTCTGATAACGAGTAATCGTTAATCCGCAAATAA
CGTAAAAACCCGCTTCGGCGGGTTTTTTTATGGGGGGAGTTTAGGGAAAGAGCATTTGTCAGAATATT
TAAGGGCGCCTGTCACTTTGCTTGATATATGAGAATTATTTAACCTTATAAATGAGAAAAAAGCAACG
CACTTTAAATAAGATACGTTGCTTTTTCGATTGATGAACACCTATAATTAAACTATTCATCTATTATTTA
TGATTTTTTGTATATACAATATTTCTAGTTTGTTAAAGAGAATTAAGAAAATAAATCTCGAAAATAATA
AAGGGAAAATCAGTTTTTGATATCAAAATTATACATGTCAACGATAATACAAAATATAATACAAACTA
TAAGATGTTATCAGTATTTATTATGCATTTAGAATACGTACTCAGCGTCTGGGTTCCCCATCGGTGATG
TCGTATAAGAGACGTATAGGAGACCTATAGTGTCTTCGGGG//CCGAATTCGCAGGTGgcagCACCTGCgat
c//TTAACCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGCATG
TGTTTTATCGGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATATTTACA
CTTACATACTAATAATAAATTCAACAAACAATTTATTTATGTTTATTTATTTATTAAAAAAAAACAAAA
ACTCAAAATTTCTTCTATAAAGTAACAAAACTTTTAGGATCTAATTCAATTAGAGACTAATTCAATTAG
AGCTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCGACCGCCGGAGTATAAATAGAGGCG
CTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAG
CAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGCTCGAAGCCGGTCGCCACCatgGCCTCCTCC
GAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCACGAGTT
CGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACC
AAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTA
CGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGC
GCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTcCAGGACGGCTCC
TTCATCTACAAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGAC
TATGGGCTGGGAGGCgTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACA
AGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGAA
GCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGACATCACCTCCCACAACGAGGACT
ACACCATCGTGGAGCAGTACGAGCGCGCCGAGGGCCGCCACCACCTGTTCCTGTAGCGGCCGCGACTC
TAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCC
CTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTAC
AAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTG
TCCAAACTCATCAATGTATCTTAGATATCTATAACAAGAAAATATATATATAATAAGTTATCACGTAA

GTAGAACATGAAATAACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATCA
TGCGTCATTTTGACTCACGCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAAGCACGC
CTCACGGGAGCTCCAAGCGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTCGCGCTATTTAGAA
AGAGAGAGCAATATTTCAAGAATGCATGCGTCAATTTTACGCAGACTATCTTTCTAGGGTTAA//gGAAG
AGCcgtcGCTCTTCC//CCGGGAAAAGAAGACCATATACGTCTCCTATAGGTCTCGTATACTGCTTCCTAT
ACGACATCACCGATGGGGAACACGAAGCGATCCAGCCCCCCTAATTAGCCCGGGCGAAAGGCCCAGT
ACGAAAGGCTCAGTCGAAAGTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCATGGGGAGTCCCCAC
ACTACCATCGGCGCTACGGCGTTTCACTTCTGAGTTCGGCATGGGGTCAGGTGGGACCACCGCGCTAC
TGCCGCCAGGCAAACAAGGGGTGTTATGAGCCATATTCAGGTATAAATGGGCTCGCGATAATGTTCAG
AATTGGTTAATTGGTTGTAACACTGACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCG
CTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAATATGAGTATTCAACA
TTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTG
GTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACA
GCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGC
TATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCT
CAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAG
AATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGA
GGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGA
ACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCGATGGCAACA
ACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGAT
GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATA
AATCCGGAGCCGGTGAGCGTGGTTCTCGCGGTATCATCGCAGCGCTGGGGCCAGATGGTAAGCCCTCC
CGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTG
AGATAGGTGCCTCACTGATTAAGCATTGGTAAAAGCAGAGCATTACGCTGACTTGACGGGACGGCGCA
AGCTCATGACCAAAATCCCTTAACGTGAGTTACGCGCGCGTCGTTCCACTGAGCGTCAGACCCCGT

Red: dsRed protein sequence

Grey: pBac vector sequence between the two homology arms that will initially be inserted into the genome but removed when the pBac is excised

Green: the paired AarI restriction enzyme sites (for 5' homology arm insertion) and the paired SapI sites (for 3' homology arm insertion)

// flanking the restriction sites: locations of AarI or SapI cuts (the short sequence between the // will be removed when the circular vector is digested with the restriction enzyme); Gibson primer overlaps should be designed to begin adjacent to these

Underlined sequences: the sequences typically included in primers to provide overlap for Gibson reactions

(A more detailed vector map [e.g. 3xP3 promoter, transposon sequences] can be provided upon request or viewed at dgrc.bio.indiana.edu, vector #1364.)

Sequencing primers (I did not name these; don't blame me for the F/R confusion):

pHD R: ACGAAAGGCTCAGTCGAAAG

Hsp70R: CGGTCGAGGGTTCGAAATCGATAAG
SV40 F: GGCCGCGACTCTAGATCATAATC
pHD F: TGATATCAAAATTATACATGTCAACG

(We have modified this basic vector to generate several variants, including ones that contain either a C-terminal 3x HA tag, a GFP tag, or a GAL4 insertion; we have also substituted various fluorescent proteins in place of the dsRed, which should enable simultaneous injection and recovery of several different genomic insertions. [To substitute dsRed with other fluorescent proteins, pBac can be digested with NotI and NcoI to remove dsRed; restriction enzyme or Gibson primers can then be used to amplify another fluorescent protein and integrate it into pBac.])

To create an N-terminally FLAG-tagged murashka

Using the standard pBac vector, we insert two homology arms, one 5' of the dsRed sequence (AarI site) and the other 3' of the dsRed sequence (SapI site). Because, after excision, the pBac sequence leaves a "TTAA", it is necessary to place the homology arms immediately adjacent to but not including the same genomic TTAA, so that while the TTAA is not included in your donor, upon excision the TTAA left by the transposase event is accommodated and the genome editing appears seamless. Therefore, in the case where we're tagging the N-terminus of murashka with a 3xFLAG tag, the donor template will include a ~1kb 5' arm, the FLAG tag, a small fragment running up to but not including the next TTAA in the genome, the dsRed and associated sequence, and the 3' arm.

The ~2500bp surrounding murashka's ATG start codon are here. This sequence will be used to design the homology arms to be inserted into the piggyBac vector.

. . .
AACGATGGGTTATTGTGTTTAACAACTTAAAATGGAAACGCTGAAAGTGTAAAATTTATCGTTACCAC
ATATCATGGCAGACGGAGTCGGCGGAGTTAGTT
CACATATACACATACATACATATGTACTACCTATGTACCCGTAACTGTTAACTTTATATCTATATCTAT
ATATATCGTATCTGTATTTGTATCGGCAGCTTTTCCACATACCCCCTATACATGAATTTATATACATTTA
TATCGTACTTATACCTATAGTTGAAAAACTAACCACGCACGTATCTCATCTCAGTTCCATATAATTGAA
AACCTATTAAAGCTACACTCTTTAAACAAACTGAAAAATTGTTGCTGTTTTATTTATTACGATTCCAAT
AAATATTTACGTATATCCACACATAAACACACCAACACGTATATATACATATCAATATATACTCACATT
TATATATATAAATTTTATTATCCTTGCCTAAGTAAGTATTTGGCCTACTAGAGAAACCATAAATGTTTT
GTTATTGAGCATGTTGGCTTGTTGGTTGTTTGCTTTCTGTTCTGGAGCGCGGGTCCGGCTTAGAATACTT
CTTCCTAAACAAATAATAATAAAAAACTCAAAAATGTAATATCAAATACCAAAAATTAAAATCCACAA
ACATCTAGAGCGTAAAAAACCGTTTTCCGTTTTCGATTATAATCTGAGTTGGTGGCATCACGAACTTGA

ACGAACTCAACTTTTTGATTAAAATACAAAATATCAGCTAATAGTTATTGTAACTAAATATACCTGATA
TGATTTATCAATTTCTCAAGAATTCGCGGAAGTAAAGTAAATTTTCTTGAGTGATTCGGGCCTAGAAAA
AAAAGCTCTTTTTAGGACAAAAGATAATGCTATACAATAATGCTTTCATTTCTACCCTTGAGTCGCCTG
TATTTGTAAAACAAAAATGTTCAATTATGATCAGTTTTGGATGGTCCCAATCGTATTCAGATTGCAAAC
GGAACCGGGATGCATAGTTAAGTGTTCAAGCGGCCACTTAATAAGCAAATCGTCTCGTTTCTAATTCC
GCCCGCAGCTGTGTGATTACAAGAAAAGCGCAAAGAAAGCTCTCGCCCAGTCGAAAGTCGAATAGCA
ATAGCAGTGGCATCCACAACAGCAGAAGCAGCAGCAACAACAACAGGCAGTAGCAACAACCGCAGG
CAGGAGCAACAACAACAGTGCGGCAGCAGCATCTGCAGCAACAACAGTGCGGCAGCAGCATGCTGAA
CTCGAATGCCACCGGCGGCAACAATCATGGGGGCAGCAGCCGTCGGCCCTTTACCCGCAATTCTTCAT
CGGTCCGCTCGCAACGCGGCGGTGGTGGCGGGCTGGGCGGTCGCATGGTATACTCGCCACACACACAT
CCGCACGTCCAGCAGCAGCAACAACAACAACAGCAGCAACAGCAACAGGGGCAGCAGCGGTCCAGC
AATAACGGCAACAACAACCTGTGGCTGGGCAACAGTCCGTGGTGCAACGTCAATTTAAACGGTGGCA
ACAGCGGCAACAACAATAACAACGGTGGCAACAACAACAATAACAACGGTGGCAACATCAGCAGCA
ACAACAACAGCAACAATGTCAATAACAAAGATGCAAACCCAAATTGGTAAGAGCCGAGAGCCGCTTG
CGTTTTTAAACCATTTCCACCTACTCACTCGAGATCCTTTCCTGGAATCGAGCTCTATGTTTAGCCTTAA
GTGACATTGAAAAGAAAAGCGAATTCGAAACGTTACCCCCAAATCCAATCCGAACCGCTCCCGAAAG
CCCCCAATGAATATCCATCCATAATTAACAAGACATATCCTTGTAACTTCCCCCACGACTAGAACAGTT
TATCTACTTTCTATCTCTATCTCTATTACACCTGAACTGAAAAAAAAAGTCAACAAGAACCGAAAATC
GCTTTTAGGCTAAAGGGAGAATTTTGTAATTAATGTCGCGTTCAGTAGTGCAACCAAAAACATTTGAA
TTGGTCCTAACCGTTGAATTTTCGAAGCACTAAACAAAAAAGCTTGCCTTGCCAATTTGATGGCGACA
CACAATGGATATAGTGTTATACCATTTATACTTCTCGATCATTTGTAACTCTCTTGCGTTATCTTGCGTA
CTGTACATTTTGTGTTTTGTTCCTTGTCGAAATTGGCACCCGAAACCCACATCCTAACCCACTCAGTTA
CCATACCATACCATTATGTACTAATTAACGTTGTAACTATCCCAGCATTGACATCAAAAACTACACATA
CACACCCACATTATACTATCCAACAAGAACTGAGTTTTGATTTGAATTTACTTAATGGTTTACCTATCG
TATGTACATTTTTTAACCGCTTTTAAATTACCATACCCGAAATTTGAGCAGTGTGAAGTTACGTAGAAG
ATATCGTATGCGTATTAACCGAATCAGAAACTCTTAATTTCCATAATATCTTGTAATAACTTGAAGAAC
ATTCAATGTTTCATAGACTTAAGACCGATCAAACAGTTCCCCAAGAAAATGCATGAAAAGCAC . . .

Underlined: the sequence used for the 5’ homology arm, using PCR primers (with Gibson overlaps in *italics*)

5’ arm F: *TAGGAGACCTATAGTGTCTTCGGGG* CTAACCACGCACGTATCTCATCTC
5’ arm R: *ctttatcatcatcatctttataatccat* GCTGCTGCCGCACTG

3’ homology arm:

3’ arm F: *CGCAGACTATCTTTCTAGGGTTAA* ACGGTGGCAACAGCGG
3’ arm R: *GACGTATATGGTCTTCTTTTCCCGG* GGGGAACTGTTTGATCGGTC

Black: TTAA fragment; the TTAA fragment is the sequence immediately after the tag site that leads right up to *but does not include* a TTAA. The 3’ arm is designed to begin immediately after, *but not include*, that same TTAA. When the pBac vector sequence is excised, it leaves behind a TTAA, which your TTAA fragment and 3’ arm accommodate.

(Because the TTAA forward primer includes the 3x FLAG tag, upon PCRing the TTAA fragment, the TTAA fragment will include Gibson overlap sequence with the upstream 5'arm, the 3xFLAG tag, the TTAA fragment itself, and Gibson overlap with the downstream pBac vector.)

TTAA frag F: *GCAGCAACAACAGTGCGGCAGCAGC*
atggattataaagatgatgatgataaagattataaagatgatgatgataaagattataaagatgatgatgataaaactagt
ATGCTGAACTCGAATGCCAC
TTAA frag R: *CAATATGATTATCTTTCTAGGGTTAA* ATTGACGTTGCACCACGG

Final order of the fragments once in the vector: (vector sequence) – 5' homology arm – 3xFLAG tag – TTAA fragment – dsRed and vector sequence – 3' homology arm – (vector sequence)

The pBac vector is digested with either AarI or SapI (order usually doesn't matter, unless the 3' arm has an AarI site or the 5' arm or TTAA fragment have a SapI site), and the homology arms and fragment are amplified and gel extracted. I have had the better success when the vector concentration is at least ~100ng/uL and the arms/fragment are more than 100ng/uL. I set up Gibson reactions with ~100ng of vector and >200ng of arms/fragment if I can. After the 1hr Gibson I digest with whichever enzyme I used to open the homology arm spot—if I'm inserting the 5' arm, I will after that [5' arm + vector] Gibson reaction digest for at least a few hours with AarI again [if the 5' arm has successfully been inserted, the restriction site will be destroyed, so digesting at this step only removes previously undigested vector]. This is in order to minimize any original vector that escaped initial digestion. I transform electrocompetent cells using 60uL water, 2-4uL cells, and ~4uL Gibson reaction DNA. (pBac vectors encode amp resistance.)

Once one arm is in and checked by sequencing, digest the vector with the other site's enzyme and insert the other arm. Once both homology arms and TTAA fragment is inserted, the full sequence of the pBac vector will be

AGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGG
CTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGCCCACCACTTCAAGA
ACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATA
AGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACG
GGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG

AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGT
CGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGG
TTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAA
CGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCG
TTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGA
ACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGGCGAGAGTAGGGAACTGCCAGGCATC
AAACTAAGCAGAAGGCCCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTCTGTGTTGTAAAAC
GACGGCCAGTCTTAAGCTCGGGCCCCCTGGGCGGTTCTGATAACGAGTAATCGTTAATCCGCAAATAA
CGTAAAAACCCGCTTCGGCGGGTTTTTTTATGGGGGGAGTTTAGGGAAAGAGCATTTGTCAGAATATT
TAAGGGCGCCTGTCACTTTGCTTGATATATGAGAATTATTTAACCTTATAAATGAGAAAAAAGCAACG
CACTTTAAATAAGATACGTTGCTTTTTCGATTGATGAACACCTATAATTAAACTATTCATCTATTATTTA
TGATTTTTTGTATATACAATATTTCTAGTTTGTTAAAGAGAATTAAGAAAATAAATCTCGAAAATAATA
AAGGGAAAATCAGTTTTTGATATCAAAATTATACATGTCAACGATAATACAAAATATAATACAAACTA
TAAGATGTTATCAGTATTTATTATGCATTTAGAATACGTACTCAGCGTCTGGGTTCCCCATCGGTGATG
TCGTATAAGAGACGTATAGGAGACCTATAGTGTCTTCGGGGCTAACCACGCACGTATCTCATCTCAGT
TCCATATAATTGAAAACCTATTAAAGCTACACTCTTTAAACAAACTGAAAAATTGTTGCTGTTTTATTT
ATTACGATTCCAATAAATATTTACGTATATCCACACATAAACACACCAACACGTATATATACATATCA
ATATATACTCACATTTATATATATAAATTTTATTATCCTTGCCTAAGTAAGTATTTGGCCTACTAGAGA
AACCATAAATGTTTTGTTATTGAGCATGTTGGCTTGTTGGTTGTTTGCTTTCTGTTCTGGAGCGCGGGTC
CGGCTTAGAATACTTCTTCCTAAACAAATAATAATAAAAAACTCAAAAATGTAATATCAAATACCAAA
AATTAAAATCCACAAACATCTAGAGCGTAAAAAACCGTTTTCCGTTTTCGATTATAATCTGAGTTGGTG
GCATCACGAACTTGAACGAACTCAACTTTTTGATTAAAATACAAAATATCAGCTAATAGTTATTGTAA
CTAAATATACCTGATATGATTTATCAATTTCTCAAGAATTCGCGGAAGTAAAGTAAATTTTCTTGAGTG
ATTCGGGCCTAGAAAAAAAAGCTCTTTTTAGGACAAAAGATAATGCTATACAATAATGCTTTCATTTCT
ACCCTTGAGTCGCCTGTATTTGTAAAACAAAAATGTTCAATTATGATCAGTTTTGGATGGTCCCAATCG
TATTCAGATTGCAAACGGAACCGGGATGCATAGTTAAGTGTTCAAGCGGCCACTTAATAAGCAAATCG
TCTCGTTTCTAATTCCGCCCGCAGCTGTGTGATTACAAGAAAAGCGCAAAGAAAGCTCTCGCCCAGTC
GAAAGTCGAATAGCAATAGCAGTGGCATCCACAACAGCAGAAGCAGCAGCAACAACAACAGGCAGT
AGCAACAACCGCAGGCAGGAGCAACAACAACAGTGCGGCAGCAGCATCTGCAGCAACAACAGTGCG
GCAGCAGCatggattataaagatgatgatgataaagattataaagatgatgatgataaagattataaagatgatgatgataaaactagATGCTGAACTCG
AATGCCACCGGCGGCAACAATCATGGGGGCAGCAGCCGTCGGCCCTTTACCCGCAATTCTTCATCGGT
CCGCTCGCAACGCGGCGGTGGTGGCGGGCTGGGCGGTCGCATGGTATACTCGCCACACACACATCCGC
ACGTCCAGCAGCAGCAACAACAACAACAGCAGCAACAGCAACAGGGGCAGCAGCGGTCCAGCAATA
ACGGCAACAACAACCTGTGGCTGGGCAACAGTCCGTGGTGCAACGTCAATTTAACCCTAGAAAGATA
ATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGCATGTGTTTTATCGGTCTGTATA
TCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATATTTACACTTACATACTAATAATAAA
TTCAACAAACAATTTATTTATGTTTATTTATTTATTAAAAAAAAACAAAAACTCAAAATTTCTTCTATA
AAGTAACAAAACTTTTAGGATCTAATTCAATTAGAGACTAATTCAATTAGAGCTAATTCAATTAGGAT
CCAAGCTTATCGATTTCGAACCCTCGACCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGAC
AATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGC
TGAACAAGCTAAACAATCGGCTCGAAGCCGGTCGCCACCatgGCCTCCTCCGAGGACGTCATCAAGGAG
TTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGG
GCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCC
CTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCG
ACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG
GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTcCAGGACGGCTCCTTCATCTACAAGGTGAA

GTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACTATGGGCTGGGAGGCgT
CCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGGCCCTGAAGCTGAA
GGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGAAGCCCGTGCAGCTGCCCG
GCTACTACTACGTGGACTCCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAG
TACGAGCGCGCCGAGGGCCGCCACCACCTGTTCCTGTAGCGGCCGCGACTCTAGATCATAATCAGCCA
TACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAA
AATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCAT
CACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGT
ATCTTAGATATCTATAACAAGAAAATATATATATAATAAGTTATCACGTAAGTAGAACATGAAATAAC
AATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATCATGCGTCATTTTGACTCAC
GCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAAGCACGCCTCACGGGAGCTCCAAG
CGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTCGCGCTATTTAGAAAGAGAGAGCAATATTTC
AAGAATGCATGCGTCAATTTTACGCAGACTATCTTTCTAGGGTTAAACGGTGGCAACAGCGGCAACAA
CAATAACAACGGTGGCAACAACAACAATAACAACGGTGGCAACATCAGCAGCAACAACAACAGCAAC
AATGTCAATAACAAAGATGCAAACCCAAATTGGTAAGAGCCGAGAGCCGCTTGCGTTTTTAAACCATT
TCCACCTACTCACTCGAGATCCTTTCCTGGAATCGAGCTCTATGTTTAGCCTTAAGTGACATTGAAAAG
AAAAGCGAATTCGAAACGTTACCCCCAAATCCAATCCGAACCGCTCCCGAAAGCCCCCAATGAATATC
CATCCATAATTAACAAGACATATCCTTGTAACTTCCCCCACGACTAGAACAGTTTATCTACTTTCTATC
TCTATCTCTATTACACCTGAACTGAAAAAAAAAGTCAACAAGAACCGAAAATCGCTTTTAGGCTAAAG
GGAGAATTTTGTAATTAATGTCGCGTTCAGTAGTGCAACCAAAAACATTTGAATTGGTCCTAACCGTTG
AATTTTCGAAGCACTAAACAAAAAAGCTTGCCTTGCCAATTTGATGGCGACACACAATGGATATAGTG
TTATACCATTTATACTTCTCGATCATTTGTAACTCTCTTGCGTTATCTTGCGTACTGTACATTTTGTGTTT
TGTTCCTTGTCGAAATTGGCACCCGAAACCCACATCCTAACCCACTCAGTTACCATACCATACCATTAT
GTACTAATTAACGTTGTAACTATCCCAGCATTGACATCAAAAACTACACATACACACCCACATTATACT
ATCCAACAAGAACTGAGTTTTGATTTGAATTTACTTAATGGTTTACCTATCGTATGTACATTTTTTAACC
GCTTTTAAATTACCATACCCGAAATTTGAGCAGTGTGAAGTTACGTAGAAGATATCGTATGCGTATTA
ACCGAATCAGAAACTCTTAATTTCCATAATATCTTGTAATAACTTGAAGAACATTCAATGTTTCATAGA
CTTAAGACCGATCAAACAGTTCCCCCCGGGAAAAGAAGACCATATACGTCTCCTATAGGTCTCGTATA
CTGCTTCCTATACGACATCACCGATGGGGAACACGAAGCGATCCAGCCCCCCTAATTAGCCCGGGCGA
AAGGCCCAGTACGAAAGGCTCAGTCGAAAGTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCATGGG
GAGTCCCCACACTACCATCGGCGCTACGGCGTTTCACTTCTGAGTTCGGCATGGGGTCAGGTGGGACC
ACCGCGCTACTGCCGCCAGGCAAACAAGGGGTGTTATGAGCCATATTCAGGTATAAATGGGCTCGCGA
TAATGTTCAGAATTGGTTAATTGGTTGTAACACTGACCCCTATTTGTTTATTTTTCTAAATACATTCAAA
TATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAATATGA
GTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCA
GAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGG
ATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTA
AAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATA
CACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGAC
AGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAA
CGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGAT
CGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCGA
TGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATA
GACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTAT
TGCTGATAAATCCGGAGCCGGTGAGCGTGGTTCTCGCGGTATCATCGCAGCGCTGGGGCCAGATGGTA
AGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAG

ATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAAAGCAGAGCATTACGCTGACTTGACGGGA
CGGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTACGCGCGCGTCGTTCCACTGAGCGTCAGAC
CCCGT

Red: dsRed protein sequence

Grey: pBac vector sequence between the two homology arms that will initially be inserted into the genome but removed when the pBac is excised

Green: the paired AarI restriction enzyme sites (for 5' homology arm insertion) and the paired SapI sites (for 3' homology arm insertion)

// flanking the restriction sites: locations of AarI or SapI cuts (the short sequence between the // will be removed when the circular vector is digested with the restriction enzyme); Gibson primer overlaps should be designed to begin adjacent to these

Underlined: 5' homology arm and 3' homology arm, using PCR primers (with Gibson overlaps in *italics*)

Blue: 3xFLAG tag immediately before start codon

Black: TTAA fragment

Once this vector is constructed, it can be injected along with U6.2B (gRNA vector) into Cas9-expressing *Drosophila* embryos. (Typically 100uL of DNA containing 500ng/uL pBac and 100ng/uL U6.2B is given to whoever is injecting, and they can inject that concentration or further dilute it.)

Picking gRNAs

For *Drosophila*, using the pU6-BbsI-chiRNA vector from the O'Connor-Giles lab has worked well for us. A detailed protocol for gRNA annealing, ligation, and double digestion can currently be found at http://flycrispr.molbio.wisc.edu/protocols/gRNA.

Choosing gRNA sequences to use with the pBac homology-directed repair can be a little challenging. You don't want to use a gRNA that recognizes any sequence in your pBac donor; if you do, the Cas9 will also cut your pBac. One way to avoid this is by mutating the PAM sequence (NGG) in your pBac, ideally so it maintains the amino acid sequence. If you're making a precise deletion, you can ideally find gRNAs within the region that your pBac is missing. If you get lucky, you can find a reasonable gRNA (i.e. without too many predicted off-target sites) that is interrupted by the pBac vector sequence in your donor—if the gRNA recognizes the genomic sequence exactly where your 5' and 3' arms meet in the

genome, your pBac sequence will break up the two halves of the gRNA target sequence. (Searching the gRNA options for “TTAA” is an easy way to do this.) In this way, your gRNA will cut the genomic locus, but will not find that same sequence in your donor vector.

For the FLAG-tagged murashka, one gRNA’s sequence was TGCAACGTCAATTTAAACGG(TGG), with the dsRed cassette falling directly between the TTAA and the rest of the gRNA.

For each gRNA target site, two oligos must be ordered, with overhangs to ligate into the BbsI-chiRNA vector as detailed at http://flycrispr.molbio.wisc.edu/protocols/gRNA:
CTTCGTGCAACGTCAATTTAAACGG and AAACCCGTTTAAATTGACGTTGCAC

Drosophila embryos are injected and left to hatch, and then all combined into a single bottle. If pBac integration into the genome is successful, the injected flies’ progeny will have two fluorescent red segments towards the larval tail, and as adults will have red fluorescent eyes. These dsRed-expressing flies (assumed to be heterozygous at this stage) can be crossed to a balancer and used to make a homozygous stock. The sequence for the integrated 3x FLAG pBac vector would be

AACGATGGGTTATTGTGTTTAACAACTTAAAATGGAAACGCTGAAAGTGTAAAATTTATCGTTACCAC
ATATCATGGCAGACGGAGTCGGCGGAGTTAGTTCACATATACACATACATACATATGTACTACCTATG
TACCCGTAACTGTTAACTTTATATCTATATCTATATATATCGTATCTGTATTTGTATCGGCAGCTTTTCC
ACATACCCCCTATACATGAATTTATATACATTTATATCGTACTTATACCTATAGTTGAAAAACTAACCA
CGCACGTATCTCATCTCAGTTCCATATAATTGAAAACCTATTAAAGCTACACTCTTTAAACAAACTGAA
AAATTGTTGCTGTTTTATTTATTACGATTCCAATAAATATTTACGTATATCCACACATAAACACACCAA
CACGTATATATACATATCAATATATACTCACATTTATATATATAAATTTTATTATCCTTGCCTAAGTAA
GTATTTGGCCTACTAGAGAAACCATAAATGTTTTGTTATTGAGCATGTTGGCTTGTTGGTTGTTTGCTTT
CTGTTCTGGAGCGCGGGTCCGGCTTAGAATACTTCTTCCTAAACAAATAATAATAAAAAACTCAAAAA
TGTAATATCAAATACCAAAAATTAAAATCCACAAACATCTAGAGCGTAAAAAACCGTTTTCCGTTTTC
GATTATAATCTGAGTTGGTGGCATCACGAACTTGAACGAACTCAACTTTTTGATTAAAATACAAAATA
TCAGCTAATAGTTATTGTAACTAAATATACCTGATATGATTTATCAATTTCTCAAGAATTCGCGGAAGT
AAAGTAAATTTTCTTGAGTGATTCGGGCCTAGAAAAAAAAGCTCTTTTTAGGACAAAAGATAATGCTA
TACAATAATGCTTTCATTTCTACCCTTGAGTCGCCTGTATTTGTAAAACAAAAATGTTCAATTATGATC
AGTTTTGGATGGTCCCAATCGTATTCAGATTGCAAACGGAACCGGGATGCATAGTTAAGTGTTCAAGC
GGCCACTTAATAAGCAAATCGTCTCGTTTCTAATTCCGCCCGCAGCTGTGTGATTACAAGAAAAGCGC
AAAGAAAGCTCTCGCCCAGTCGAAAGTCGAATAGCAATAGCAGTGGCATCCACAACAGCAGAAGCAG
CAGCAACAACAACAGGCAGTAGCAACAACCGCAGGCAGGAGCAACAACAACAGTGCGGCAGCAGCA
TCTGCAGCAACAACAGTGCGGCAGCAGCatggattataaagatgatgatgataaagattataaagatgatgatgataaagattataaagatgatg
atgataaaactagATGCTGAACTCGAATGCCACCGGCGGCAACAATCATGGGGGCAGCAGCCGTCGGCCCTTT
ACCCGCAATTCTTCATCGGTCCGCTCGCAACGCGGCGGTGGTGGCGGGCTGGGCGGTCGCATGGTATA
CTCGCCACACACACATCCGCACGTCCAGCAGCAGCAACAACAACAACAGCAGCAACAGCAACAGGGG

CAGCAGCGGTCCAGCAATAACGGCAACAACAACCTGTGGCTGGGCAACAGTCCGTGGTGCAACGTCA
ATTTAACCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGCAT
GTGTTTTATCGGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATATTTAC
ACTTACATACTAATAATAAATTCAACAAACAATTTATTTATGTTTATTTATTTATTAAAAAAAAACAAA
AACTCAAAATTTCTTCTATAAAGTAACAAAACTTTTAGGATCTAATTCAATTAGAGACTAATTCAATTA
GAGCTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCGACCGCCGGAGTATAAATAGAGGC
GCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAA
GCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGCTCGAAGCCGGTCGCCACCatgGCCTCCTC
CGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCACGAG
TTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGA
CCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGT
ACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAG
CGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTcCAGGACGGCTC
CTTCATCTACAAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGA
CTATGGGCTGGGAGGCgTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCAC
AAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGA
AGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGACATCACCTCCCACAACGAGGAC
TACACCATCGTGGAGCAGTACGAGCGCGCCGAGGGCCGCCACCACCTGTTCCTGTAGCGGCCGCGACT
CTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCC
CCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTA
CAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTT
GTCCAAACTCATCAATGTATCTTAGATATCTATAACAAGAAAATATATATATAATAAGTTATCACGTA
AGTAGAACATGAAATAACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATC
ATGCGTCATTTTGACTCACGCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAAGCACG
CCTCACGGGAGCTCCAAGCGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTCGCGCTATTTAGA
AAGAGAGAGCAATATTTCAAGAATGCATGCGTCAATTTTACGCAGACTATCTTTCTAGGGTTAAACGG
TGGCAACAGCGGCAACAACAATAACAACGGTGGCAACAACAACAATAACAACGGTGGCAACATCAGC
AGCAACAACAACAGCAACAATGTCAATAACAAAGATGCAAACCCAAATTGGTAAGAGCCGAGAGCCG
CTTGCGTTTTTAAACCATTTCCACCTACTCACTCGAGATCCTTTCCTGGAATCGAGCTCTATGTTTAGCC
TTAAGTGACATTGAAAAGAAAAGCGAATTCGAAACGTTACCCCCAAATCCAATCCGAACCGCTCCCGA
AAGCCCCCAATGAATATCCATCCATAATTAACAAGACATATCCTTGTAACTTCCCCCACGACTAGAAC
AGTTTATCTACTTTCTATCTCTATCTCTATTACACCTGAACTGAAAAAAAAAAGTCAACAAGAACCGAAA
ATCGCTTTTAGGCTAAAGGGAGAATTTTGTAATTAATGTCGCGTTCAGTAGTGCAACCAAAAACATTT
GAATTGGTCCTAACCGTTGAATTTTCGAAGCACTAAACAAAAAAGCTTGCCTTGCCAATTTGATGGCG
ACACACAATGGATATAGTGTTATACCATTTATACTTCTCGATCATTTGTAACTCTCTTGCGTTATCTTGC
GTACTGTACATTTTGTGTTTTGTTCCTTGTCGAAATTGGCACCCGAAACCCACATCCTAACCCACTCAG
TTACCATACCATACCATTATGTACTAATTAACGTTGTAACTATCCCAGCATTGACATCAAAAACTACAC
ATACACACCCACATTATACTATCCAACAAGAACTGAGTTTTGATTTGAATTTACTTAATGGTTTACCTA
TCGTATGTACATTTTTTAACCGCTTTTAAATTACCATACCCGAAATTTGAGCAGTGTGAAGTTACGTAG
AAGATATCGTATGCGTATTAACCGAATCAGAAACTCTTAATTTCCATAATATCTTGTAATAACTTGAAG
AACATTCAATGTTTCATAGACTTAAGACCGATCAAACAGTTCCCCAAGAAAATGCATGAAAAGCAC

This fly is then crossed to a pBac transposase fly (e.g. Bloomington Stock #32070, #32071, or #32073), and it’s when the progeny have one copy of the pBac dsRed and one copy of the transposase that excision occurs. These progeny are crossed to balancers, and single flies that no longer express dsRed (and do not

have the pBac transposase chromosome, usually marked by CFP eyes) are used to set up single fly crosses with the balancer line.

N.B. To remove dsRed, I find it easiest to use a transposase line on whichever chromosome your dsRed sits. If FLAGmura dsRed is on 3rd chromosome, it's easiest to cross that fly with the third chromosome transposase. When this heterozygous fly is crossed with a balancer, the progeny will have either a FLAGmura chromosome over a balancer or a transposase chromosome over a balancer. You then know you can disregard any fly with a transposase chromosome (marked by CFP eyes), and the remaining flies—the ones with FLAGmura and hopefully some now missing dsRed—can be checked for dsRed expression. Usually about 10% of flies will have no detectable dsRed expression in the eye, and these are used in single fly crosses to set up lines where FLAGmura is now (potentially) cleanly integrated in the genome. The majority of flies I've sequenced at this point have the dsRed removed perfectly. (Once or twice I've had issues with a modified chromosome not going homozygous, but re-removing the dsRed has solved this problem.)

After excision of the dsRed cassette, the scarless 3x FLAG tag sequence will be

. . .
AACGATGGGTTATTGTGTTTAACAACTTAAAATGGAAACGCTGAAAGTGTAAAATTTATCGTTACCAC
ATATCATGGCAGACGGAGTCGGCGGAGTTAGTTCACATATACACATACATACATATGTACTACCTATG
TACCCGTAACTGTTAACTTTATATCTATATCTATATATATCGTATCTGTATTTGTATCGGCAGCTTTTCC
ACATACCCCCTATACATGAATTTATATACATTTATATCGTACTTATACCTATAGTTGAAAAACTAACCA
CGCACGTATCTCATCTCAGTTCCATATAATTGAAAACCTATTAAAGCTACACTCTTTAAACAAACTGAA
AAATTGTTGCTGTTTTATTTATTACGATTCCAATAAATATTTACGTATATCCACACATAAACACACCAA
CACGTATATATACATATCAATATATACTCACATTTATATATATAAATTTTATTATCCTTGCCTAAGTAA
GTATTTGGCCTACTAGAGAAACCATAAATGTTTTGTTATTGAGCATGTTGGCTTGTTGGTTGTTTGCTTT
CTGTTCTGGAGCGCGGGTCCGGCTTAGAATACTTCTTCCTAAACAAATAATAATAAAAAACTCAAAAA
TGTAATATCAAATACCAAAAATTAAAATCCACAAACATCTAGAGCGTAAAAAACCGTTTTCCGTTTTC
GATTATAATCTGAGTTGGTGGCATCACGAACTTGAACGAACTCAACTTTTTGATTAAAATACAAAATA
TCAGCTAATAGTTATTGTAACTAAATATACCTGATATGATTTATCAATTTCTCAAGAATTCGCGGAAGT
AAAGTAAATTTTCTTGAGTGATTCGGGCCTAGAAAAAAAAGCTCTTTTTAGGACAAAAGATAATGCTA
TACAATAATGCTTTCATTTCTACCCTTGAGTCGCCTGTATTTGTAAAACAAAAATGTTCAATTATGATC
AGTTTTGGATGGTCCCAATCGTATTCAGATTGCAAACGGAACCGGGATGCATAGTTAAGTGTTCAAGC
GGCCACTTAATAAGCAAATCGTCTCGTTTCTAATTCCGCCCGCAGCTGTGTGATTACAAGAAAAGCGC
AAAGAAAGCTCTCGCCCAGTCGAAAGTCGAATAGCAATAGCAGTGGCATCCACAACAGCAGAAGCAG
CAGCAACAACAACAGGCAGTAGCAACAACCGCAGGCAGGAGCAACAACAACAGTGCGGCAGCAGCA
TCTGCAGCAACAACAGTGCGGCAGCAGCatggattataaagatgatgatgataaagattataaagatgatgatgataaagattataaagatgatg
atgataaaactagATGCTGAACTCGAATGCCACCGGCGGCAACAATCATGGGGGCAGCAGCCGTCGGCCCTTT
ACCCGCAATTCTTCATCGGTCCGCTCGCAACGCGGCGGTGGTGGCGGGCTGGGCGGTCGCATGGTATA
CTCGCCACACACACATCCGCACGTCCAGCAGCAGCAACAACAACAACAGCAGCAACAGCAACAGGGG
CAGCAGCGGTCCAGCAATAACGGCAACAACAACCTGTGGCTGGGCAACAGTCCGTGGTGCAACGTCA

ATTTAAACGGTGGCAACAGCGGCAACAACAATAACAACGGTGGCAACAACAACAATAACAACGGTGG
CAACATCAGCAGCAACAACAACAGCAACAATGTCAATAACAAAGATGCAAACCCAAATTGGTAAGAG
CCGAGAGCCGCTTGCGTTTTTAAACCATTTCCACCTACTCACTCGAGATCCTTTCCTGGAATCGAGCTC
TATGTTTAGCCTTAAGTGACATTGAAAAGAAAAGCGAATTCGAAACGTTACCCCCAAATCCAATCCGA
ACCGCTCCCGAAAGCCCCCAATGAATATCCATCCATAATTAACAAGACATATCCTTGTAACTTCCCCCA
CGACTAGAACAGTTTATCTACTTTCTATCTCTATCTCTATTACACCTGAACTGAAAAAAAAAGTCAACA
AGAACCGAAAATCGCTTTTAGGCTAAAGGGAGAATTTTGTAATTAATGTCGCGTTCAGTAGTGCAACC
AAAAACATTTGAATTGGTCCTAACCGTTGAATTTTCGAAGCACTAAACAAAAAAGCTTGCCTTGCCAA
TTTGATGGCGACACACAATGGATATAGTGTTATACCATTTATACTTCTCGATCATTTGTAACTCTCTTGC
GTTATCTTGCGTACTGTACATTTTGTGTTTTGTTCCTTGTCGAAATTGGCACCCGAAACCCACATCCTAA
CCCACTCAGTTACCATACCATACCATTATGTACTAATTAACGTTGTAACTATCCCAGCATTGACATCAA
AAACTACACATACACACCCACATTATACTATCCAACAAGAACTGAGTTTTGATTTGAATTTACTTAATG
GTTTACCTATCGTATGTACATTTTTTAACCGCTTTTAAATTACCATACCCGAAATTTGAGCAGTGTGAA
GTTACGTAGAAGATATCGTATGCGTATTAACCGAATCAGAAACTCTTAATTTCCATAATATCTTGTAAT
AACTTGAAGAACATTCAATGTTTCATAGACTTAAGACCGATCAAACAGTTCCCCAAGAAAATGCATGA
AAAGCAC . . .

At any point in this process flies can be checked by sequencing PCR products, but sequencing is especially important when the dsRed cassette is thought to be integrated and then again when it is excised.

This same strategy can be used for C-terminal tags (though leaving an extra TTAA immediately after the tag's stop codon is less of a concern, so the TTAA fragment may not be necessary), GAL4 insertions immediately following the start codon, inserting unique protease cleavage sites at certain locations within the protein, altering splice site sequences, creating specific in-frame deletions, and much else.

To insert a GAL4 sequence into the endogenous locus

The GAL4 system in *Drosophila* allows controlled expression of specific transgenes only in those tissues where the GAL4 is expressed. Expression of the GAL4 is dictated by the context of the genome—ideally, if a GAL4 sequence is inserted into a gene encoding a sugar receptor, only in those neurons that express that sugar receptor will the GAL4 be made. The purpose of the GAL4 is to bind to the UAS sequence and promote expression of whatever transgene has been placed downstream of the UAS: to identify the cells' location, UAS-GFP; to silence neuronal activity constitutively, UAS-Kir2.1, an inward-rectifying potassium channel; to activate neurons using red-shifted light, UAS-ReaChR, a red-light responsive nonselective cation channel, and so forth. If inserted into the endogenous locus, with a stop codon at the end of the GAL4 sequence (and perhaps some of the first coding exon sequence removed), the GAL4 insertion can also eliminate that protein's expression.

Using the same *murashka* genomic area as above, the 5’ arm for the GAL4 pBac will be
5’ arm F new: *TAGGAGACCTATAGTGTCTTCGGGG* cgtcgaatgaagaagtgatcgcgaaaagtgacgcc
5’ arm R new: *GTTCGATAGAAGACAGTAGCTTCAT* GCTGCTGCCGCACTGTTG

CGTCGAATGAAGAAGTGATCGCGAAAAGTGACGCCTATTTGAAAGCAAAGGAGAAATCCTACTACAT
AAATGGTATTGAAAAATTGAAGCGAACTTTGTTGAATGAAAAACTTAAATTTGGCAGAAAAAAAAAT
GTTTACTGTGGTAGGTCGGGGACTTTTTAATGGACCTGTTATGTATATGTGGGAGGGGCGGCTTAACG
ATGGGTTATTGTGTTTAACAACTTAAAATGGAAACGCTGAAAGTGTAAAATTTATCGTTACCACATATC
ATGGCAGACGGAGTCGGCGGAGTTAGTTCACATATACACATACATACATATGTACTACCTATGTACCC
GTAACTGTTAACTTTATATCTATATCTATATATATCGTATCTGTATTTGTATCGGCAGCTTTTCCACATA
CCCCCTATACATGAATTTATATACATTTATATCGTACTTATACCTATAGTTGAAAAACTAACCACGCAC
GTATCTCATCTCAGTTCCATATAATTGAAAACCTATTAAAGCTACACTCTTTAAACAAACTGAAAAATT
GTTGCTGTTTTATTTATTACGATTCCAATAAATATTTACGTATATCCACACATAAACACACCAACACGT
ATATATACATATCAATATATACTCACATTTATATATATAAATTTTATTATCCTTGCCTAAGTAAGTATTT
GGCCTACTAGAGAAACCATAAATGTTTTGTTATTGAGCATGTTGGCTTGTTGGTTGTTTGCTTTCTGTTC
TGGAGCGCGGGTCCGGCTTAGAATACTTCTTCCTAAACAAATAATAATAAAAAACTCAAAAATGTAAT
ATCAAATACCAAAAATTAAAATCCACAAACATCTAGAGCGTAAAAAACCGTTTTCCGTTTTCGATTAT
AATCTGAGTTGGTGGCATCACGAACTTGAACGAACTCAACTTTTTGATTAAAATACAAAATATCAGCT
AATAGTTATTGTAACTAAATATACCTGATATGATTTATCAATTTCTCAAGAATTCGCGGAAGTAAAGTA
AATTTTCTTGAGTGATTCGGGCCTAGAAAAAAAAAGCTCTTTTTAGGACAAAAGATAATGCTATACAAT
AATGCTTTCATTTCTACCCTTGAGTCGCCTGTATTTGTAAAACAAAAATGTTCAATTATGATCAGTTTTG
GATGGTCCCAATCGTATTCAGATTGCAAACGGAACCGGGATGCATAGTTAAGTGTTCAAGCGGCCACT
TAATAAGCAAATCGTCTCGTTTCTAATTCCGCCCGCAGCTGTGTGATTACAAGAAAAGCGCAAAGAAA
GCTCTCGCCCAGTCGAAAGTCGAATAGCAATAGCAGTGGCATCCACAACAGCAGAAGCAGCAGCAAC
AACAACAGGCAGTAGCAACAACCGCAGGCAGGAGCAACAACAACAGTGCGGCAGCAGCATCTGCAG
CAACAACAGTGCGGCAGCAGC

The GAL4 is amplified by the following primers, with overlaps to the 5’ arm and pBac vector:
GAL4 F primer: ATGAAGCTACTGTCTTCTATCGAAC
GAL4 R primer: *CAATATGATTATCTTTCTAGGGTTAA* TTAGTTAGTTACTCTTTTTTTGGGTTTG

ATGAAGCTACTGTCTTCTATCGAACAAGCATGCGATATTTGCCGACTTAAAAAGCTCAAGTGCTCCAA
AGAAAAACCGAAGTGCGCCAAGTGTCTGAAGAACAACTGGGAGTGTCGCTACTCTCCCAAAACCAAA
AGGTCTCCGCTGACTAGGGCACATCTGACAGAAGTGGAATCAAGGCTAGAAAGACTGGAACAGCTAT
TTCTACTGATTTTTCCTCGAGAAGACCTTGACATGATTTTGAAAATGGATTCTTTACAGGATATAAAAG
CATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAAAGATGCCGTCACAGATAGATTGGCTTCA
GTGGAGACTGATATGCCTCTAACATTGAGACAGCATAGAATAAGTGCGACATCATCATCGGAAGAGA
GTAGTAACAAAGGTCAAAGACAGTTGACTGTATCGATTGACTCGGCAGCTCATCATGATAACTCCACA
ATTCCGTTGGATTTTATGCCCAGGGATGCTCTTCATGGATTTGATTGGTCTGAAGAGGATGACATGTCG
GATGGCTTGCCCTTCCTGAAAACGGACCCCAACAATAATGGGTTCTTTGGCGACGGTTCTCTCTTATGT
ATTCTTCGATCTATTGGCTTTAAACCGGAAAATTACACGAACTCTAACGTTAACAGGCTCCCGACCATG
ATTACGGATAGATACACGTTGGCTTCTAGATCCACAACATCCCGTTTACTTCAAAGTTATCTCAATAAT
TTTCACCCCTACTGCCCTATCGTGCACTCACCGACGCTAATGATGTTGTATAATAACCAGATTGAAATC
GCGTCGAAGGATCAATGGCAAATCCTTTTTAACTGCATATTAGCCATTGGAGCCTGGTGTATAGAGGG

GGAATCTACTGATATAGATGTTTTTTACTATCAAAATGCTAAATCTCATTTGACGAGCAAGGTCTTCGA
GTCAGGTTCCATAATTTTGGTGACAGCCCTACATCTTCTGTCGCGATATACACAGTGGAGGCAGAAAA
CAAATACTAGCTATAATTTTCACAGCTTTTCCATAAGAATGGCCATATCATTGGGCTTGAATAGGGACC
TCCCCTCGTCCTTCAGTGATAGCAGCATTCTGGAACAAAGACGCCGAATTTGGTGGTCTGTCTACTCTT
GGGAGATCCAATTGTCCCTGCTTTATGGTCGATCCATCCAGCTTTCTCAGAATACAATCTCCTTCCCTTC
TTCTGTCGACGATGTGCAGCGTACCACAACAGGTCCCACCATATATCATGGCATCATTGAAACAGCAA
GGCTCTTACAAGTTTTCACAAAAATCTATGAACTAGACAAAACAGTAACTGCAGAAAAAAGTCCTATA
TGTGCAAAAAAATGCTTGATGATTTGTAATGAGATTGAGGAGGTTTCGAGACAGGCACCAAAGTTTTT
ACAAATGGATATTTCCACCACCGCTCTAACCAATTTGTTGAAGGAACACCCTTGGCTATCCTTTACAAG
ATTCGAACTGAAGTGGAAACAGTTGTCTCTTATCATTTATGTATTAAGAGATTTTTTCACTAATTTTACC
CAGAAAAAGTCACAACTAGAACAGGATCAAAATGATCATCAAAGTTATGAAGTTAAACGATGCTCCA
TCATGTTAAGCGATGCAGCACAAAGAACTGTTATGTCTGTAAGTAGCTATATGGACAATCATAATGTC
ACCCCATATTTTGCCTGGAATTGTTCTTATTACTTGTTCAATGCAGTCCTAGTACCCATAAAGACTCTAC
TCTCAAACTCAAAATCGAATGCTGAGAATAACGAGACCGCACAATTATTACAACAAATTAACACTGTT
CTGATGCTATTAAAAAAACTGGCCACTTTTAAAATCCAGACTTGTGAAAAATACATTCAAGTACTGGA
AGAGGTATGTGCGCCGTTTCTGTTATCACAGTGTGCAATCCCATTACCGCATATCAGTTATAACAATAG
TAATGGTAGCGCCATTAAAAATATTGTCGGTTCTGCAACTATCGCCCAATACCCTACTCTTCCGGAGGA
AAATGTCAACAATATCAGTGTTAAATATGTTTCTCCTGGCTCAGTAGGGCCTTCACCTGTGCCATTGAA
ATCAGGAGCAAGTTTCAGTGATCTAGTCAAGCTGTTATCTAACCGTCCACCCTCTCGTAACTCTCCAGT
GACAATACCAAGAAGCACACCTTCGCATCGCTCAGTCACGCCTTTTCTAGGGCAACAGCAACAGCTGC
AATCATTAGTGCCACTGACCCCGTCTGCTTTGTTTGGTGGCGCCAATTTTAATCAAAGTGGGAATATTG
CTGATAGCTCATTGTCCTTCACTTTCACTAACAGTAGCAACGGTCCGAACCTCATAACAACTCAAACAA
ATTCTCAAGCGCTTTCACAACCAATTGCCTCCTCTAACGTTCATGATAACTTCATGAATAATGAAATCA
CGGCTAGTAAAATTGATGATGGTAATAATTCAAAACCACTGTCACCTGGTTGGACGGACCAAACTGCG
TATAACGCGTTTGGAATCACTACAGGGATGTTTAATACCACTACAATGGATGATGTATATAACTATCTA
TTCGATGATGAAGATACCCCACCAAACCCAAAAAAAGAGTAACTAACTAA

The 3’ homology arm is amplified by the following primers:

3’ arm F: *TAGGAGACCTATAGTGTCTTCGGGG* TCATCGGTCCGCTCGCAACG

3’ arm R: *GACGTATATGGTCTTCTTTTCCCGG* CAAGAAGAAGCGTTGGGGACA

CTGAACTCGAATGCCACCGGCGGCAACAATCATGGGGGCAGCAGCCGTCGGCCCTTTACCCGCAATTC
TTCATCGGTCCGCTCGCAACGCGGCGGTGGTGGCGGGCTGGGCGGTCGCATGGTATACTCGCCACACA
CACATCCGCACGTCCAGCAGCAGCAACAACAACAACAGCAGCAACAGCAACAGGGGCAGCAGCGGTC
CAGCAATAACGGCAACAACAACCTGTGGCTGGGCAACAGTCCGTGGTGCAACGTCAATTTAAACGGT
GGCAACAGCGGCAACAACAATAACAACGGTGGCAACAACAACAATAACAACGGTGGCAACATCAGC
AGCAACAACAACAGCAACAATGTCAATAACAAAGATGCAAACCCAAATTGGTAAGAGCCGAGAGCCG
CTTGCGTTTTTAAACCATTTCCACCTACTCACTCGAGATCCTTTCCTGGAATCGAGCTCTATGTTTAGCC
TTAAGTGACATTGAAAAGAAAAGCGAATTCGAAACGTTACCCCCAAATCCAATCCGAACCGCTCCCGA
AAGCCCCCAATGAATATCCATCCATAATTAACAAGACATATCCTTGTAACTTCCCCCACGACTAGAAC
AGTTTATCTACTTTCTATCTCTATCTCTATTACACCTGAACTGAAAAAAAAAGTCAACAAGAACCGAAA
ATCGCTTTTAGGCTAAAGGGAGAATTTTGTAATTAATGTCGCGTTCAGTAGTGCAACCAAAAACATTT
GAATTGGTCCTAACCGTTGAATTTTCGAAGCACTAAACAAAAAAGCTTGCCTTGCCAATTTGATGGCG
ACACACAATGGATATAGTGTTATACCATTTATACTTCTCGATCATTTGTAACTCTCTTGCGTTATCTTGC
GTACTGTACATTTTGTGTTTTGTTCCTTGTCGAAATTGGCACCCGAAACCCACATCCTAACCCACTCAG
TTACCATACCATACCATTATGTACTAATTAACGTTGTAACTATCCCAGCATTGACATCAAAAACTACAC

ATACACACCCACATTATACTATCCAACAAGAACTGAGTTTTGATTTGAATTTACTTAATGGTTTACCTA
TCGTATGTACATTTTTTAACCGCTTTTAAATTACCATACCCGAAATTTGAGCAGTGTGAAGTTACGTAG
AAGATATCGTATGCGTATTAACCGAATCAGAAACTCTTAATTTCCATAATATCTTGTAATAACTTGAAG
AACATTCAATGTTTCATAGACTTAAGACCGATCAAACAGTTCCCCAAGAAAATGCATGAAAAGCACTT
GAAAATGGCGTACCAAACCTTATTACGTAGCAATTTCTCTTTCGTTTCGTAGTTGTTTCTTTGTCCCCAA
CGCTTCTTCTTG

The fully assembled muraGAL4 pBac vector sequence is

AGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGG
CTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGCCCACCACTTCAAGA
ACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATA
AGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACG
GGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGT
CGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGG
TTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAA
CGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCG
TTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGA
ACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGGCGAGAGTAGGGAACTGCCAGGCATC
AAACTAAGCAGAAGGCCCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTCTGTGTTGTAAAAC
GACGGCCAGTCTTAAGCTCGGGCCCCCTGGGCGGTTCTGATAACGAGTAATCGTTAATCCGCAAATAA
CGTAAAAACCCGCTTCGGCGGGTTTTTTTATGGGGGGAGTTTAGGGAAAGAGCATTTGTCAGAATATT
TAAGGGCGCCTGTCACTTTGCTTGATATATGAGAATTATTTAACCTTATAAATGAGAAAAAAGCAACG
CACTTTAAATAAGATACGTTGCTTTTTCGATTGATGAACACCTATAATTAAACTATTCATCTATTATTTA
TGATTTTTTGTATATACAATATTTCTAGTTTGTTAAAGAGAATTAAGAAAATAAATCTCGAAAATAATA
AAGGGAAAATCAGTTTTTGATATCAAAATTATACATGTCAACGATAATACAAAATATAATACAAACTA
TAAGATGTTATCAGTATTTATTATGCATTTAGAATACGTACTCAGCGTCTGGGTTCCCCATCGGTGATG
TCGTATAAGAGACGTATAGGAGACCTATAGTGTCTTCGGGG

CAGAATTTGTTAGCTTTCGGTCCATATCCAAATCCTCAGGTCTCAAAGGAATGCTCGCTGGGAAGAAA
TTAACGTCGAATGAAGAAGTGATCGCGAAAAGTGACGCCTATTTGAAAGCAAAGGAGAAATCCTACT
ACATAAATGGTATTGAAAAATTGAAGCGAACTTTGTTGAATGAAAAACTTAAATTTGGCAGAAAAAA
AAATGTTTACTGTGGTAGGTCGGGGACTTTTTAATGGACCTGTTATGTATATGTGGGAGGGGCGGCTTA
ACGATGGGTTATTGTGTTTAACAACTTAAAATGGAAACGCTGAAAGTGTAAAATTTATCGTTACCACA
TATCATGGCAGACGGAGTCGGCGGAGTTAGTTCACATATACACATACATACATATGTACTACCTATGT
ACCCGTAACTGTTAACTTTATATCTATATCTATATATATCGTATCTGTATTTGTATCGGCAGCTTTTCCA
CATACCCCCTATACATGAATTTATATACATTTATATCGTACTTATACCTATAGTTGAAAAACTAACCAC
GCACGTATCTCATCTCAGTTCCATATAATTGAAAACCTATTAAAGCTACACTCTTTAAACAAACTGAAA
AATTGTTGCTGTTTTATTTATTACGATTCCAATAAATATTTACGTATATCCACACATAAACACACCAAC
ACGTATATATACATATCAATATATACTCACATTTATATATATAAATTTTATTATCCTTGCCTAAGTAAGT
ATTTGGCCTACTAGAGAAACCATAAATGTTTTGTTATTGAGCATGTTGGCTTGTTGGTTGTTTGCTTTCT
GTTCTGGAGCGCGGGTCCGGCTTAGAATACTTCTTCCTAAACAAATAATAATAAAAAACTCAAAAATG
TAATATCAAATACCAAAAATTAAAATCCACAAACATCTAGAGCGTAAAAAACCGTTTTCCGTTTTCGA

TTATAATCTGAGTTGGTGGCATCACGAACTTGAACGAACTCAACTTTTTGATTAAAATACAAAATATCA
GCTAATAGTTATTGTAACTAAATATACCTGATATGATTTATCAATTTCTCAAGAATTCGCGGAAGTAAA
GTAAATTTTCTTGAGTGATTCGGGCCTAGAAAAAAAAGCTCTTTTTAGGACAAAAGATAATGCTATAC
AATAATGCTTTCATTTCTACCCTTGAGTCGCCTGTATTTGTAAAACAAAAATGTTCAATTATGATCAGT
TTTGGATGGTCCCAATCGTATTCAGATTGCAAACGGAACCGGGATGCATAGTTAAGTGTTCAAGCGGC
CACTTAATAAGCAAATCGTCTCGTTTCTAATTCCGCCCGCAGCTGTGTGATTACAAGAAAAGCGCAAA
GAAAGCTCTCGCCCAGTCGAAAGTCGAATAGCAATAGCAGTGGCATCCACAACAGCAGAAGCAGCAG
CAACAACAACAGGCAGTAGCAACAACCGCAGGCAGGAGCAACAACAACAGTGCGGCAGCAGCATCT
GCAGCAACAACAGTGCGGCAGCAGC

ATGAAGCTACTGTCTTCTATCGAACAAGCATGCGATATTTGCCGACTTAAAAAGCTCAAGTGCTCCAA
AGAAAAACCGAAGTGCGCCAAGTGTCTGAAGAACAACTGGGAGTGTCGCTACTCTCCCAAAACCAAA
AGGTCTCCGCTGACTAGGGCACATCTGACAGAAGTGGAATCAAGGCTAGAAAGACTGGAACAGCTAT
TTCTACTGATTTTTCCTCGAGAAGACCTTGACATGATTTTGAAAATGGATTCTTTACAGGATATAAAAG
CATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAAAGATGCCGTCACAGATAGATTGGCTTCA
GTGGAGACTGATATGCCTCTAACATTGAGACAGCATAGAATAAGTGCGACATCATCATCGGAAGAGA
GTAGTAACAAAGGTCAAAGACAGTTGACTGTATCGATTGACTCGGCAGCTCATCATGATAACTCCACA
ATTCCGTTGGATTTTATGCCCAGGGATGCTCTTCATGGATTTGATTGGTCTGAAGAGGATGACATGTCG
GATGGCTTGCCCTTCCTGAAAACGGACCCCAACAATAATGGGTTCTTTGGCGACGGTTCTCTCTTATGT
ATTCTTCGATCTATTGGCTTTAAACCGGAAAATTACACGAACTCTAACGTTAACAGGCTCCCGACCATG
ATTACGGATAGATACACGTTGGCTTCTAGATCCACAACATCCCGTTTACTTCAAAGTTATCTCAATAAT
TTTCACCCCTACTGCCCTATCGTGCACTCACCGACGCTAATGATGTTGTATAATAACCAGATTGAAATC
GCGTCGAAGGATCAATGGCAAATCCTTTTTAACTGCATATTAGCCATTGGAGCCTGGTGTATAGAGGG
GGAATCTACTGATATAGATGTTTTTTACTATCAAAATGCTAAATCTCATTTGACGAGCAAGGTCTTCGA
GTCAGGTTCCATAATTTTGGTGACAGCCCTACATCTTCTGTCGCGATATACACAGTGGAGGCAGAAAA
CAAATACTAGCTATAATTTTCACAGCTTTTCCATAAGAATGGCCATATCATTGGGCTTGAATAGGGACC
TCCCCTCGTCCTTCAGTGATAGCAGCATTCTGGAACAAAGACGCCGAATTTGGTGGTCTGTCTACTCTT
GGGAGATCCAATTGTCCCTGCTTTATGGTCGATCCATCCAGCTTTCTCAGAATACAATCTCCTTCCCTTC
TTCTGTCGACGATGTGCAGCGTACCACAACAGGTCCCACCATATATCATGGCATCATTGAAACAGCAA
GGCTCTTACAAGTTTTCACAAAAATCTATGAACTAGACAAAACAGTAACTGCAGAAAAAAGTCCTATA
TGTGCAAAAAAATGCTTGATGATTTGTAATGAGATTGAGGAGGTTTCGAGACAGGCACCAAAGTTTTT
ACAAATGGATATTTCCACCACCGCTCTAACCAATTTGTTGAAGGAACACCCTTGGCTATCCTTTACAAG
ATTCGAACTGAAGTGGAAACAGTTGTCTCTTATCATTTATGTATTAAGAGATTTTTTCACTAATTTTACC
CAGAAAAAGTCACAACTAGAACAGGATCAAAATGATCATCAAAGTTATGAAGTTAAACGATGCTCCA
TCATGTTAAGCGATGCAGCACAAAGAACTGTTATGTCTGTAAGTAGCTATATGGACAATCATAATGTC
ACCCCATATTTTGCCTGGAATTGTTCTTATTACTTGTTCAATGCAGTCCTAGTACCCATAAAGACTCTAC
TCTCAAACTCAAAATCGAATGCTGAGAATAACGAGACCGCACAATTATTACAACAAATTAACACTGTT
CTGATGCTATTAAAAAAACTGGCCACTTTTAAAATCCAGACTTGTGAAAAATACATTCAAGTACTGGA
AGAGGTATGTGCGCCGTTTCTGTTATCACAGTGTGCAATCCCATTACCGCATATCAGTTATAACAATAG
TAATGGTAGCGCCATTAAAAATATTGTCGGTTCTGCAACTATCGCCCAATACCCTACTCTTCCGGAGGA
AAATGTCAACAATATCAGTGTTAAATATGTTTCTCCTGGCTCAGTAGGGCCTTCACCTGTGCCATTGAA
ATCAGGAGCAAGTTTCAGTGATCTAGTCAAGCTGTTATCTAACCGTCCACCCTCTCGTAACTCTCCAGT
GACAATACCAAGAAGCACACCTTCGCATCGCTCAGTCACGCCTTTTCTAGGGCAACAGCAACAGCTGC
AATCATTAGTGCCACTGACCCCGTCTGCTTTGTTTGGTGGCGCCAATTTTAATCAAAGTGGGAATATTG
CTGATAGCTCATTGTCCTTCACTTTCACTAACAGTAGCAACGGTCCGAACCTCATAACAACTCAAACAA
ATTCTCAAGCGCTTTCACAACCAATTGCCTCCTCTAACGTTCATGATAACTTCATGAATAATGAAATCA

CGGCTAGTAAAATTGATGATGGTAATAATTCAAAACCACTGTCACCTGGTTGGACGGACCAAACTGCG
TATAACGCGTTTGGAATCACTACAGGGATGTTTAATACCACTACAATGGATGATGTATATAACTATCTA
TTCGATGATGAAGATACCCCACCAAACCCAAAAAAAGAGTAACTAACTAA

TTAACCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGCATGT
GTTTTATCGGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATATTTACAC
TTACATACTAATAATAAATTCAACAAACAATTTATTTATGTTTATTTATTTATTAAAAAAAAACAAAAA
CTCAAAATTTCTTCTATAAAGTAACAAAACTTTTAGGATCTAATTCAATTAGAGACTAATTCAATTAGA
GCTAATTCAATTAGGATCCAAGCTTATCGATTTCGAACCCTCGACCGCCGGAGTATAAATAGAGGCGC
TTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGC
AAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGCTCGAAGCCGGTCGCCACCatgGCCTCCTCCG
AGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCACGAGTTC
GAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCA
AGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTAC
GTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCG
CGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTcCAGGACGGCTCCT
TCATCTACAAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACT
ATGGGCTGGGAGGCgTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAA
GGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGAAG
CCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGACATCACCTCCCACAACGAGGACTA
CACCATCGTGGAGCAGTACGAGCGCGCCGAGGGCCGCCACCACCTGTTCCTGTAGCGGCCGCGACTCT
AGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCC
TGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACA
AATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGT
CCAAACTCATCAATGTATCTTAGATATCTATAACAAGAAAATATATATATAATAAGTTATCACGTAAG
TAGAACATGAAATAACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATCAT
GCGTCATTTTGACTCACGCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAAGCACGCC
TCACGGGAGCTCCAAGCGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTCGCGCTATTTAGAAA
GAGAGAGCAATATTTCAAGAATGCATGCGTCAATTTTACGCAGACTATCTTTCTAGGGTTAA

CTGAACTCGAATGCCACCGGCGGCAACAATCATGGGGGCAGCAGCCGTCGGCCCTTTACCCGCAATTC
TTCATCGGTCCGCTCGCAACGCGGCGGTGGTGGCGGGCTGGGCGGTCGCATGGTATACTCGCCACACA
CACATCCGCACGTCCAGCAGCAGCAACAACAACAACAGCAGCAACAGCAACAGGGGCAGCAGCGGTC
CAGCAATAACGGCAACAACAACCTGTGGCTGGGCAACAGTCCGTGGTGCAACGTCAATTTAAACGGT
GGCAACAGCGGCAACAACAATAACAACGGTGGCAACAACAACAATAACAACGGTGGCAACATCAGC
AGCAACAACAACAGCAACAATGTCAATAACAAAGATGCAAACCCAAATTGGTAAGAGCCGAGAGCCG
CTTGCGTTTTTAAACCATTTCCACCTACTCACTCGAGATCCTTTCCTGGAATCGAGCTCTATGTTTAGCC
TTAAGTGACATTGAAAAGAAAAGCGAATTCGAAACGTTACCCCCAAATCCAATCCGAACCGCTCCCGA
AAGCCCCCAATGAATATCCATCCATAATTAACAAGACATATCCTTGTAACTTCCCCCACGACTAGAAC
AGTTTATCTACTTTCTATCTCTATCTCTATTACACCTGAACTGAAAAAAAAAGTCAACAAGAACCGAAA
ATCGCTTTTAGGCTAAAGGGAGAATTTTGTAATTAATGTCGCGTTCAGTAGTGCAACCAAAAACATTT
GAATTGGTCCTAACCGTTGAATTTTCGAAGCACTAAACAAAAAAGCTTGCCTTGCCAATTTGATGGCG
ACACACAATGGATATAGTGTTATACCATTTATACTTCTCGATCATTTGTAACTCTCTTGCGTTATCTTGC
GTACTGTACATTTTGTGTTTTGTTCCTTGTCGAAATTGGCACCCGAAACCCACATCCTAACCCACTCAG
TTACCATACCATACCATTATGTACTAATTAACGTTGTAACTATCCCAGCATTGACATCAAAAACTACAC
ATACACACCCACATTATACTATCCAACAAGAACTGAGTTTTGATTTGAATTTACTTAATGGTTTACCTA

TCGTATGTACATTTTTTAACCGCTTTTAAATTACCATACCCGAAATTTGAGCAGTGTGAAGTTACGTAG
AAGATATCGTATGCGTATTAACCGAATCAGAAACTCTTAATTTCCATAATATCTTGTAATAACTTGAAG
AACATTCAATGTTTCATAGACTTAAGACCGATCAAACAGTTCCCCAAGAAAATGCATGAAAAGCACTT
GAAAATGGCGTACCAAACCTTATTACGTAGCAATTTCTCTTTCGTTTCGTAGTTGTTTCTTTGTCCCCAA
CGCTTCTTCTTG

CCGGGAAAAGAAGACCATATACGTCTCCTATAGGTCTCGTATACTGCTTCCTATACGACATCACCGAT
GGGGAACACGAAGCGATCCAGCCCCCCTAATTAGCCCGGGCGAAAGGCCCAGTCTTTCGACTGAGCCT
TTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCATGGGGAGTCCCCACACTACCATCGGCGCTA
CGGCGTTTCACTTCTGAGTTCGGCATGGGGTCAGGTGGGACCACCGCGCTACTGCCGCCAGGCAAACA
AGGGGTGTTATGAGCCATATTCAGGTATAAATGGGCTCGCGATAATGTTCAGAATTGGTTAATTGGTT
GTAACACTGACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAA
CCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAATATGAGTATTCAACATTTCCGTGTCGCCCTT
ATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGAT
GCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGA
GAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATT
ATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTG
AGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCC
ATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAA
CCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAA
GCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCGATGGCAACAACGTTGCGCAAACTATT
AACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTG
CAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCCGGAGCCGGTGAGC
GTGGTTCTCGCGGTATCATCGCAGCGCTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACA
CGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGAT
TAAGCATTGGTAAAAGCAGAGCATTACGCTGACTTGACGGGACGGCGCAAGCTCATGACCAAAATCC
CTTAACGTGAGTTACGCGCGCGTCGTTCCACTGAGCGTCAGACCCCGT

The 5’ and 3’ homology arms are underlined, and the GAL4 sequence is highlighted in grey.

Because the GAL4 sequence is replacing a large part of the first coding exon and will create a null allele, introducing an extra TTAA immediately after the GAL4’s stop codon is not a problem and no TTAA fragment was engineered to accommodate it.

References

Ahmed, A.E., and Labavitch, J.M. (1980). Cell Wall Metabolism in Ripening Fruit: II. CHANGES IN CARBOHYDRATE-DEGRADING ENZYMES IN RIPENING: BARTLETT' PEARS. Plant Physiol *65*, 1014-1016.

Akalal, D.B., Yu, D., and Davis, R.L. (2011). The long-term memory trace formed in the Drosophila alpha/beta mushroom body neurons is abolished in long-term memory mutants. J Neurosci *31*, 5643-5647.

Alberini, C.M. (2009). Transcription factors in long-term memory and synaptic plasticity. Physiol Rev *89*, 121-145.

Alberini, C.M., Ghirardi, M., Metz, R., and Kandel, E.R. (1994). C/EBP is an immediate-early gene required for the consolidation of long-term facilitation in Aplysia. Cell *76*, 1099-1114.

Alberti, S., Halfmann, R., and Lindquist, S. (2010). Biochemical, cell biological, and genetic assays to analyze amyloid and prion aggregation in yeast. Methods Enzymol *470*, 709-734.

Anselme, P., and Robinson, M.J. (2016). "Wanting," "liking," and their relation to consciousness. J Exp Psychol Anim Learn Cogn *42*, 123-140.

Aristotle (350 BC). On Memory and Reminiscence.

Aso, Y., Sitaraman, D., Ichinose, T., Kaun, K.R., Vogt, K., Belliart-Guerin, G., Placais, P.Y., Robie, A.A., Yamagata, N., Schnaitmann, C.*, et al.* (2014). Mushroom body output neurons encode valence and guide memory-based action selection in Drosophila. Elife *3*, e04580.

Awasaki, T., and Kimura, K. (1997). pox-neuro is required for development of chemosensory bristles in Drosophila. J Neurobiol *32*, 707-721.

Bailey, C.H., Kandel, E.R., and Si, K. (2004). The persistence of long-term memory: a molecular approach to self-sustaining changes in learning-induced synaptic growth. Neuron *44*, 49-57.

Baines, R.A., Uhler, J.P., Thompson, A., Sweeney, S.T., and Bate, M. (2001). Altered electrical properties in Drosophila neurons developing without synaptic transmission. J Neurosci *21*, 1523-1531.

Barretto, R.P., Gillis-Smith, S., Chandrashekar, J., Yarmolinsky, D.A., Schnitzer, M.J., Ryba, N.J., and Zuker, C.S. (2015). The neural representation of taste quality at the periphery. Nature *517*, 373-376.

Benito, E., and Barco, A. (2010). CREB's control of intrinsic and synaptic plasticity: implications for CREB-dependent memory models. Trends Neurosci *33*, 230-240.

Berridge, K.C. (2007). The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology (Berl) *191*, 391-431.

Berridge, K.C., and Kringelbach, M.L. (2015). Pleasure systems in the brain. Neuron *86*, 646-664.

Berridge, K.C., and Robinson, T.E. (2003). Parsing reward. Trends Neurosci *26*, 507-513.

Berridge, K.C., Robinson, T.E., and Aldridge, J.W. (2009). Dissecting components of reward: 'liking', 'wanting', and learning. Curr Opin Pharmacol *9*, 65-73.

Berry, J.A., Cervantes-Sandoval, I., Nicholas, E.P., and Davis, R.L. (2012). Dopamine is required for learning and forgetting in Drosophila. Neuron *74*, 530-542.

Bliss, T.V., and Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J Physiol *232*, 331-356.

Bourtchuladze, R., Frenguelli, B., Blendy, J., Cioffi, D., Schutz, G., and Silva, A.J. (1994). Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. Cell *79*, 59-68.

Brummel, T., Ching, A., Seroude, L., Simon, A.F., and Benzer, S. (2004). Drosophila lifespan enhancement by exogenous bacteria. Proceedings of the National Academy of Sciences of the United States of America *101*, 12974-12979.

Buetow, L., and Huang, D.T. (2016). Structural insights into the catalysis and regulation of E3 ubiquitin ligases. Nat Rev Mol Cell Biol *17*, 626-642.

Burke, C.J., Huetteroth, W., Owald, D., Perisse, E., Krashes, M.J., Das, G., Gohl, D., Silies, M., Certel, S., and Waddell, S. (2012). Layered reward signalling through octopamine and dopamine in Drosophila. Nature *492*, 433-437.

Burke, C.J., and Waddell, S. (2011). Remembering nutrient quality of sugar in Drosophila. Curr Biol *21*, 746-750.

Byers, D., Davis, R.L., and Kiger, J.A., Jr. (1981). Defect in cyclic AMP phosphodiesterase due to the dunce mutation of learning in Drosophila melanogaster. Nature *289*, 79-81.

Castellucci, V., and Kandel, E.R. (1976). Presynaptic facilitation as a mechanism for behavioral sensitization in Aplysia. Science *194*, 1176-1178.

Castellucci, V., Pinsker, H., Kupfermann, I., and Kandel, E.R. (1970). Neuronal mechanisms of habituation and dishabituation of the gill-withdrawal reflex in Aplysia. Science *167*, 1745-1748.

Cervantes-Sandoval, I., and Davis, R.L. (2012). Distinct traces for appetitive versus aversive olfactory memories in DPM neurons of Drosophila. Curr Biol *22*, 1247-1252.

Charlu, S., Wisotsky, Z., Medina, A., and Dahanukar, A. (2013). Acid sensing by sweet and bitter taste neurons in Drosophila melanogaster. Nat Commun *4*, 2042.

Chen, T.W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., *et al.* (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295-300.

Clark, D.A., Freifeld, L., and Clandinin, T.R. (2013). Mapping and cracking sensorimotor circuits in genetic model organisms. Neuron *78*, 583-595.

Colomb, J., Kaiser, L., Chabaud, M.A., and Preat, T. (2009). Parametric and genetic analysis of Drosophila appetitive long-term memory and sugar motivation. Genes Brain Behav *8*, 407-415.

Crick, F. (1984). Memory and molecular turnover. Nature *312*, 101.

Dahanukar, A., Foster, K., van der Goes van Naters, W.M., and Carlson, J.R. (2001). A Gr receptor is required for response to the sugar trehalose in taste neurons of Drosophila. Nat Neurosci *4*, 1182-1186.

Dahanukar, A., Lei, Y.T., Kwon, J.Y., and Carlson, J.R. (2007). Two Gr genes underlie sugar reception in Drosophila. Neuron *56*, 503-516.

Dash, P.K., Hochner, B., and Kandel, E.R. (1990). Injection of the cAMP-responsive element into the nucleus of Aplysia sensory neurons blocks long-term facilitation. Nature *345*, 718-721.

Dash, P.K., and Moore, A.N. (1996). Characterization and phosphorylation of CREB-like proteins in Aplysia central nervous system. Brain Res Mol Brain Res *39*, 43-51.

Davis, R.L., and Kiger, J.A., Jr. (1981). Dunce mutants of Drosophila melanogaster: mutants defective in the cyclic AMP phosphodiesterase enzyme system. J Cell Biol *90*, 101-107.

Decker, C.J., Teixeira, D., and Parker, R. (2007). Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in Saccharomyces cerevisiae. J Cell Biol *179*, 437-449.

DeFelipe, J. (2006). Brain plasticity and mental processes: Cajal again. Nat Rev Neurosci *7*, 811-817.

Deisseroth, K. (2015). Optogenetics: 10 years of microbial opsins in neuroscience. Nat Neurosci *18*, 1213-1225.

Deshaies, R.J., and Joazeiro, C.A. (2009). RING domain E3 ubiquitin ligases. Annu Rev Biochem *78*, 399-434.

Dethier, V.G. (1976). The Hungry Fly: A physiological study of the behavior associated with feeding (Cambridge, Mass.: Harvard University Press).

Dick, A.J., and Labavitch, J.M. (1989). Cell Wall Metabolism in Ripening Fruit: IV. Characterization of the Pectic Polysaccharides Solubilized during Softening of ;Bartlett' Pear Fruit. Plant Physiol *89*, 1394-1400.

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

Drachman, D.A. (2005). Do we have brain to spare? Neurology *64*, 2004-2005.

Dubnau, J., Chiang, A.S., Grady, L., Barditch, J., Gossweiler, S., McNeil, J., Smith, P., Buldoc, F., Scott, R., Certa, U.*, et al.* (2003). The staufen/pumilio pathway is involved in Drosophila long-term memory. Current biology : CB *13*, 286-296.

Dubnau, J., and Tully, T. (1998). Gene discovery in Drosophila: new insights for learning and memory. Annu Rev Neurosci *21*, 407-444.

Dudai, Y., Jan, Y.N., Byers, D., Quinn, W.G., and Benzer, S. (1976). dunce, a mutant of Drosophila deficient in learning. Proc Natl Acad Sci U S A *73*, 1684-1688.

Dunipace, L., Meister, S., McNealy, C., and Amrein, H. (2001). Spatially restricted expression of candidate taste receptors in the Drosophila gustatory system. Curr Biol *11*, 822-835.

Dus, M., Min, S., Keene, A.C., Lee, G.Y., and Suh, G.S. (2011). Taste-independent detection of the caloric content of sugar in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *108*, 11644-11649.

Freeman, E.G., and Dahanukar, A. (2015). Molecular neurobiology of Drosophila taste. Curr Opin Neurobiol *34*, 140-148.

Freeman, E.G., Wisotsky, Z., and Dahanukar, A. (2014). Detection of sweet tastants by a conserved group of insect gustatory receptors. Proceedings of the National Academy of Sciences of the United States of America *111*, 1598-1603.

French, A., Ali Agha, M., Mitra, A., Yanagawa, A., Sellier, M.J., and Marion-Poll, F. (2015). Drosophila Bitter Taste(s). Frontiers in integrative neuroscience *9*, 58.

Frey, U., and Morris, R.G. (1997). Synaptic tagging and long-term potentiation. Nature *385*, 533-536.

Fujii, S., Yavuz, A., Slone, J., Jagge, C., Song, X., and Amrein, H. (2015). Drosophila sugar receptors in sweet taste perception, olfaction, and internal nutrient sensing. Curr Biol *25*, 621-627.

Fujita, M., and Tanimura, T. (2011). Drosophila evaluates and learns the nutritional value of sugars. Curr Biol *21*, 751-755.

Glanzman, D.L. (2010). Common mechanisms of synaptic plasticity in vertebrates and invertebrates. Curr Biol *20*, R31-36.

Goldman, R., Smet, I., Singh, M. (2000). Psychiatry: Behavioral Science and Clinical Essentials (Philadelphia, PA: W.B. Saunders Company).

Gordesky-Gold, B., Rivers, N., Ahmed, O.M., and Breslin, P.A. (2008). Drosophila melanogaster prefers compounds perceived sweet by humans. Chem Senses *33*, 301-309.

Gratz, S.J., Rubinstein, C.D., Harrison, M.M., Wildonger, J., and O'Connor-Giles, K.M. (2015). CRISPR-Cas9 Genome Editing in Drosophila. Curr Protoc Mol Biol *111*, 31 32 31-20.

Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M., and O'Connor-Giles, K.M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics *196*, 961-971.

Guven-Ozkan, T., and Davis, R.L. (2014). Functional neuroanatomy of Drosophila olfactory memory formation. Learn Mem *21*, 519-526.

Hake, L.E., and Richter, J.D. (1994). CPEB is a specificity factor that mediates cytoplasmic polyadenylation during Xenopus oocyte maturation. Cell *79*, 617-627.

Hamada, F.N., Rosenzweig, M., Kang, K., Pulver, S.R., Ghezzi, A., Jegla, T.J., and Garrity, P.A. (2008). An internal thermal sensor controlling temperature preference in Drosophila. Nature *454*, 217-220.

Han, J.H., Kushner, S.A., Yiu, A.P., Cole, C.J., Matynia, A., Brown, R.A., Neve, R.L., Guzowski, J.F., Silva, A.J., and Josselyn, S.A. (2007). Neuronal competition and selection during memory formation. Science *316*, 457-460.

Harholt, J., Suttangkakul, A., and Vibe Scheller, H. (2010). Biosynthesis of pectin. Plant physiology *153*, 384-395.

Harris, D.T., Kallman, B.R., Mullaney, B.C., and Scott, K. (2015). Representations of Taste Modality in the Drosophila Brain. Neuron *86*, 1449-1460.

Hassett, C.C., Dethier, V.G., and Gans, J. (1950). A comparison of nutritive values and taste thresholds of carbohydrates for the blowfly. Biol Bull *99*, 446-453.

Hou, F., Sun, L., Zheng, H., Skaug, B., Jiang, Q.X., and Chen, Z.J. (2011). MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. Cell *146*, 448-461.

Hu, H. (2016). Reward and Aversion. Annu Rev Neurosci *39*, 297-324.

Huetteroth, W., Perisse, E., Lin, S., Klappenbach, M., Burke, C., and Waddell, S. (2015). Sweet taste and nutrient value subdivide rewarding dopaminergic neurons in Drosophila. Curr Biol *25*, 751-758.

Inagaki, H.K., Jung, Y., Hoopfer, E.D., Wong, A.M., Mishra, N., Lin, J.Y., Tsien, R.Y., and Anderson, D.J. (2014). Optogenetic control of Drosophila using a red-shifted channelrhodopsin reveals experience-dependent influences on courtship. Nat Methods *11*, 325-332.

James, W. (1890). The Principles Of Psychology (New York: Henry Holt And Company).

Jiao, Y., Moon, S.J., and Montell, C. (2007). A Drosophila gustatory receptor required for the responses to sucrose, glucose, and maltose identified by mRNA tagging. Proceedings of the National Academy of Sciences of the United States of America *104*, 14110-14115.

Jiao, Y., Moon, S.J., Wang, X., Ren, Q., and Montell, C. (2008). Gr64f is required in combination with other gustatory receptors for sugar detection in Drosophila. Current biology : CB *18*, 1797-1801.

Joseph, R.M., and Carlson, J.R. (2015). Drosophila Chemoreceptors: A Molecular Interface Between the Chemical World and the Brain. Trends in genetics : TIG *31*, 683-695.

Kain, P., and Dahanukar, A. (2015). Secondary taste neurons that convey sweet taste and starvation in the Drosophila brain. Neuron *85*, 819-832.

Kandel, E.R. (2001). The molecular biology of memory storage: a dialogue between genes and synapses. Science *294*, 1030-1038.

Kandel, E.R. (2012). The molecular biology of memory: cAMP, PKA, CRE, CREB-1, CREB-2, and CPEB. Mol Brain *5*, 14.

Kandel, E.R., Dudai, Y., and Mayford, M.R. (2014). The molecular and systems biology of memory. Cell *157*, 163-186.

Kandel, E.R., and Spencer, W.A. (1968). Cellular neurophysiological approaches in the study of learning. Physiol Rev *48*, 65-134.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J.*, et al.* (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753-767.

Khan, M.R., Li, L., Perez-Sanchez, C., Saraf, A., Florens, L., Slaughter, B.D., Unruh, J.R., and Si, K. (2015). Amyloidogenic Oligomerization Transforms Drosophila Orb2 from a Translation Repressor to an Activator. Cell *163*, 1468-1483.

Kirkhart, C., and Scott, K. (2015). Gustatory learning and processing in the Drosophila mushroom bodies. J Neurosci *35*, 5950-5958.

Kofinas, D. (2015). A tumor stole every memory I had. This is what happened when it all came back. (Quartz).

Krashes, M.J., DasGupta, S., Vreede, A., White, B., Armstrong, J.D., and Waddell, S. (2009). A neural circuit mechanism integrating motivational state with memory expression in Drosophila. Cell *139*, 416-427.

Krashes, M.J., and Waddell, S. (2011). Drosophila appetitive olfactory conditioning. Cold Spring Harb Protoc *2011*, pdb prot5609.

Kruttner, S., Stepien, B., Noordermeer, J.N., Mommaas, M.A., Mechtler, K., Dickson, B.J., and Keleman, K. (2012). Drosophila CPEB Orb2A mediates memory independent of Its RNA-binding domain. Neuron *76*, 383-395.

Lashley, K.S. (1950). In Search Of The Engram. Symposia of the Society for Experimental Biology *4*, 454-482.

LeDue, E.E., Chen, Y.C., Jung, A.Y., Dahanukar, A., and Gordon, M.D. (2015). Pharyngeal sense organs drive robust sugar consumption in Drosophila. Nature communications *6*, 6667.

Lee, Y.S., and Silva, A.J. (2009). The molecular and cellular biology of enhanced cognition. Nat Rev Neurosci *10*, 126-140.

Lima, S.Q., and Miesenbock, G. (2005). Remote control of behavior through genetically targeted photostimulation of neurons. Cell *121*, 141-152.

Liman, E.R., Zhang, Y.V., and Montell, C. (2014). Peripheral coding of taste. Neuron *81*, 984-1000.

Lisman, J. (2003). Long-term potentiation: outstanding questions and attempted synthesis. Philos Trans R Soc Lond B Biol Sci *358*, 829-842.

Liu, C., Placais, P.Y., Yamagata, N., Pfeiffer, B.D., Aso, Y., Friedrich, A.B., Siwanowicz, I., Rubin, G.M., Preat, T., and Tanimoto, H. (2012a). A subset of dopamine neurons signals reward for odour memory in Drosophila. Nature *488*, 512-516.

Liu, X., Ramirez, S., Pang, P.T., Puryear, C.B., Govindarajan, A., Deisseroth, K., and Tonegawa, S. (2012b). Optogenetic stimulation of a hippocampal engram activates fear memory recall. Nature *484*, 381-385.

Mabb, A.M., and Ehlers, M.D. (2010). Ubiquitination in postsynaptic function and plasticity. Annu Rev Cell Dev Biol *26*, 179-210.

Mabb, A.M., Je, H.S., Wall, M.J., Robinson, C.G., Larsen, R.S., Qiang, Y., Correa, S.A., and Ehlers, M.D. (2014). Triad3A regulates synaptic strength by ubiquitination of Arc. Neuron *82*, 1299-1316.

Majumdar, A., Cesario, W.C., White-Grindley, E., Jiang, H., Ren, F., Khan, M.R., Li, L., Choi, E.M., Kannan, K., Guo, F., *et al.* (2012). Critical role of amyloid-like oligomers of Drosophila Orb2 in the persistence of memory. Cell *148*, 515-529.

Markow, T.A., and O'Grady, P.M. (2005). Evolutionary genetics of reproductive behavior in Drosophila: connecting the dots. Annu Rev Genet *39*, 263-291.

Martin, K.C., Casadio, A., Zhu, H., Yaping, E., Rose, J.C., Chen, M., Bailey, C.H., and Kandel, E.R. (1997). Synapse-specific, long-term facilitation of aplysia sensory to motor synapses: a function for local protein synthesis in memory storage. Cell *91*, 927-938.

Martin, S.J., Grimwood, P.D., and Morris, R.G. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis. Annu Rev Neurosci *23*, 649-711.

Masek, P., and Scott, K. (2010). Limited taste discrimination in Drosophila. Proc Natl Acad Sci U S A *107*, 14833-14838.

Mastushita-Sakai, T., White-Grindley, E., Samuelson, J., Seidel, C., and Si, K. (2010). Drosophila Orb2 targets genes involved in neuronal growth, synapse formation, and protein turnover. Proceedings of the National Academy of Sciences of the United States of America *107*, 11987-11992.

Miesenbock, G. (2009). The optogenetic catechism. Science *326*, 395-399.

Milner, B., Squire, L.R., and Kandel, E.R. (1998). Cognitive neuroscience and the study of memory. Neuron *20*, 445-468.

Miyamoto, T., and Amrein, H. (2014). Diverse roles for the Drosophila fructose sensor Gr43a. Fly (Austin) *8*, 19-25.

Miyamoto, T., Chen, Y., Slone, J., and Amrein, H. (2013). Identification of a Drosophila glucose receptor using Ca2+ imaging of single chemosensory neurons. PloS one *8*, e56304.

Miyamoto, T., Slone, J., Song, X., and Amrein, H. (2012). A fructose receptor functions as a nutrient sensor in the Drosophila brain. Cell *151*, 1113-1125.

Miyazaki, T., Lin, T.Y., Ito, K., Lee, C.H., and Stopfer, M. (2015). A gustatory second-order neuron that connects sucrose-sensitive primary neurons and a distinct region of the gnathal ganglion in the Drosophila brain. J Neurogenet *29*, 144-155.

Montell, C. (2009). A taste of the Drosophila gustatory receptors. Curr Opin Neurobiol *19*, 345-353.

Moser, E.I., Krobert, K.A., Moser, M.B., and Morris, R.G. (1998). Impaired spatial learning after saturation of long-term potentiation. Science *281*, 2038-2042.

Musso, P.Y., Tchenio, P., and Preat, T. (2015). Delayed dopamine signaling of energy level builds appetitive long-term memory in Drosophila. Cell Rep *10*, 1023-1031.

O'Doherty, J.P., Cockburn, J., and Pauli, W.M. (2017). Learning, Reward, and Decision Making. Annu Rev Psychol *68*, 73-100.

Park, J.H., and Kwon, J.Y. (2011). A systematic analysis of Drosophila gustatory receptor gene expression in abdominal neurons which project to the central nervous system. Mol Cells *32*, 375-381.

Paul, K.R., Hendrich, C.G., Waechter, A., Harman, M.R., and Ross, E.D. (2015). Generating new prions by targeted mutation or segment duplication. Proc Natl Acad Sci U S A *112*, 8584-8589.

Penfield, W., and Perot, P. (1963). The Brain's Record of Auditory and Visual Experience. A Final Summary and Discussion. Brain *86*, 595-696.

Preston, A.R., and Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. Curr Biol *23*, R764-773.

Pulver, S.R., Pashkovski, S.L., Hornstein, N.J., Garrity, P.A., and Griffith, L.C. (2009). Temporal dynamics of neuronal activation by Channelrhodopsin-2 and TRPA1 determine behavioral output in Drosophila larvae. J Neurophysiol *101*, 3075-3088.

Ramon y Cajal, S. (1892). El nuevo concepto de la histologia de los centros nerviosos. Rev Ciencias Med Barcelona *18*, 361-541.

Ridley, E.V., Wong, A.C., and Douglas, A.E. (2013). Microbe-dependent and nonspecific effects of procedures to eliminate the resident microbiota from Drosophila melanogaster. Applied and environmental microbiology *79*, 3209-3214.

Rogan, M.T., and LeDoux, J.E. (1995). LTP is accompanied by commensurate enhancement of auditory-evoked responses in a fear conditioning circuit. Neuron *15*, 127-136.

Sacktor, T.C. (2011). How does PKMzeta maintain long-term memory? Nat Rev Neurosci *12*, 9-15.

Sato, K., Tanaka, K., and Touhara, K. (2011). Sugar-regulated cation channel formed by an insect gustatory receptor. Proc Natl Acad Sci U S A *108*, 11680-11685.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. Annu Rev Psychol *57*, 87-115.

Schwaerzel, M., Monastirioti, M., Scholz, H., Friggi-Grelin, F., Birman, S., and Heisenberg, M. (2003). Dopamine and octopamine differentiate between aversive and appetitive olfactory memories in Drosophila. The Journal of neuroscience : the official journal of the Society for Neuroscience *23*, 10495-10502.

Scott, K., Brady, R., Jr., Cravchik, A., Morozov, P., Rzhetsky, A., Zuker, C., and Axel, R. (2001). A chemosensory gene family encoding candidate gustatory and olfactory receptors in Drosophila. Cell *104*, 661-673.

Scoville, W.B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. J Neurol Neurosurg Psychiatry *20*, 11-21.

Seeger-Armbruster, S., Bosch-Bouju, C., Little, S.T., Smither, R.A., Hughes, S.M., Hyland, B.I., and Parr-Brownlie, L.C. (2015). Patterned, but not tonic, optogenetic stimulation in motor thalamus improves reaching in acute drug-induced Parkinsonian rats. J Neurosci *35*, 1211-1216.

Shepherd, J.D., and Huganir, R.L. (2007). The cell biology of synaptic plasticity: AMPA receptor trafficking. Annu Rev Cell Dev Biol *23*, 613-643.

Si, K., Giustetto, M., Etkin, A., Hsu, R., Janisiewicz, A.M., Miniaci, M.C., Kim, J.H., Zhu, H., and Kandel, E.R. (2003a). A neuronal isoform of CPEB regulates local protein synthesis and stabilizes synapse-specific long-term facilitation in aplysia. Cell *115*, 893-904.

Si, K., Lindquist, S., and Kandel, E.R. (2003b). A neuronal isoform of the aplysia CPEB has prion-like properties. Cell *115*, 879-891.

Squire, L.R., and Wixted, J.T. (2011). The cognitive neuroscience of human memory since H.M. Annu Rev Neurosci *34*, 259-288.

Stafford, J.W., Lynd, K.M., Jung, A.Y., and Gordon, M.D. (2012). Integration of taste and calorie sensing in Drosophila. J Neurosci *32*, 14767-14774.

Stevens, C.F. (1998). A million dollar question: does LTP = memory? Neuron *20*, 1-2.

Sultan, R., Stampas, A., Goldberg, M.B., and Baker, N.E. (2001). Drug resistance of bacteria commensal with Drosophila melanogaster in laboratory cultures. . In Drosophila Information Service (Drosophila Information Service), pp. 175-180.

Tempel, B.L., Bonini, N., Dawson, D.R., and Quinn, W.G. (1983). Reward learning in normal and mutant Drosophila. Proc Natl Acad Sci U S A *80*, 1482-1486.

Thoma, V., Knapek, S., Arai, S., Hartl, M., Kohsaka, H., Sirigrivatanawong, P., Abe, A., Hashimoto, K., and Tanimoto, H. (2016). Functional dissociation in sweet taste receptor neurons between and within taste organs of Drosophila. Nature communications *7*, 10678.

Tonegawa, S., Liu, X., Ramirez, S., and Redondo, R. (2015). Memory Engram Cells Have Come of Age. Neuron *87*, 918-931.

Urban, D.J., and Roth, B.L. (2015). DREADDs (designer receptors exclusively activated by designer drugs): chemogenetic tools with therapeutic utility. Annu Rev Pharmacol Toxicol *55*, 399-417.

Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet *13*, 227-232.

Vogt, K., Schnaitmann, C., Dylla, K.V., Knapek, S., Aso, Y., Rubin, G.M., and Tanimoto, H. (2014). Shared mushroom body circuits underlie visual and olfactory memories in Drosophila. Elife *3*, e02395.

Vosshall, L.B., and Stocker, R.F. (2007). Molecular architecture of smell and taste in Drosophila. Annu Rev Neurosci *30*, 505-533.

Watanabe, S., Kodaki, T., and Makino, K. (2006). Cloning, expression, and characterization of bacterial L-arabinose 1-dehydrogenase involved in an alternative pathway of L-arabinose metabolism. The Journal of biological chemistry *281*, 2612-2623.

White-Grindley, E., Li, L., Mohammad Khan, R., Ren, F., Saraf, A., Florens, L., and Si, K. (2014). Contribution of Orb2A stability in regulated amyloid-like oligomerization of Drosophila Orb2. PLoS Biol *12*, e1001786.

Wise, R.A. (2002). Brain reward circuitry: insights from unsensed incentives. Neuron *36*, 229-240.

Wisotsky, Z., Medina, A., Freeman, E., and Dahanukar, A. (2011). Evolutionary differences in food preference rely on Gr64e, a receptor for glycerol. Nature neuroscience *14*, 1534-1541.

Wu, H., and Fuxreiter, M. (2016). The Structure and Dynamics of Higher-Order Assemblies: Amyloids, Signalosomes, and Granules. Cell *165*, 1055-1066.

Yamada, T., Yang, Y., and Bonni, A. (2013). Spatial organization of ubiquitin ligase pathways orchestrates neuronal connectivity. Trends Neurosci *36*, 218-226.

Yamagata, N., Ichinose, T., Aso, Y., Placais, P.Y., Friedrich, A.B., Sima, R.J., Preat, T., Rubin, G.M., and Tanimoto, H. (2015). Distinct dopamine neurons mediate reward signals for short- and long-term memories. Proc Natl Acad Sci U S A *112*, 578-583.

Yarmolinsky, D.A., Zuker, C.S., and Ryba, N.J. (2009). Common sense about taste: from mammals to insects. Cell *139*, 234-244.

Yavuz, A., Jagge, C., Slone, J., and Amrein, H. (2014). A genetic tool kit for cellular and behavioral analyses of insect sugar receptors. Fly (Austin) *8*, 189-196.

Yin, J.C., Wallach, J.S., Del Vecchio, M., Wilder, E.L., Zhou, H., Quinn, W.G., and Tully, T. (1994). Induction of a dominant negative CREB transgene specifically blocks long-term memory in Drosophila. Cell *79*, 49-58.