Synaptic Reentry Reinforcement Based Network Model for Long-Term Memory Consolidation

Gayle M. Wittenberg, Megan R. Sullivan, and Joe Z. Tsien*

Department of Molecular Biology, Princeton University, Princeton, New Jersey

ABSTRACT: The conversion of newly formed declarative memories into long-term memories is known to be dependent on the hippocampus. Recent experiments suggest that memory consolidation requires reactivation of the NMDA receptor in CA1 during the initial week(s) after training. This led to the hypothesis that the repeated post-learning reinforcement of synaptic modifications, termed synaptic reentry reinforcement (SRR), is essential for long-term memory consolidation and storage. Based on experimental observations, we have built a computational model to further illustrate and explore the effect of the SRR process on the formation of long-term memory. We show that SRR is capable of strengthening and maintaining memory traces despite inherent variability in the system due to such processes as the turnover of synaptic receptors and their associated signaling and structural proteins. Furthermore, we demonstrate that new rounds of synaptic modification triggered by memory reactivation, either during conscious recall or sleep, could lead to the selective consolidation of a subset of memory traces. Finally, we show why the SRR process in the hippocampus is required during the initial post-training weeks for synaptic reinforcement based memory consolidation in the cortex. *Hippocampus 2002;12:637–647.* © 2002 Wiley-Liss, Inc.

KEY WORDS: SRR; long-term potentiation; NMDA receptor; hippocampus; reactivation; sleep; retrieval

INTRODUCTION

For more than a century, researchers have sought to understand the nature of memory consolidation, a crucial process underlying the formation of long-term memory (Muller and Pilzecker, 1900; McGaugh, 2000). A role for the hippocampus in long-term memory formation was first established in studies of patients suffering from hippocampal lesion (Scoville and Milner, 1957). Memory deficits have also been observed in patients with more restricted damage to the CA1 subregion of the hippocampus (Zola-Morgan et al., 1986; Rempel-Clower et al., 1996). In these studies, patients exhibit severe anterograde amnesia, as well as graded retrograde amnesia, primarily affecting recent memories. Lesion studies in animals provide further support for the temporal requirement of the hippocampus in the process of long-term memory formation (Kim and Fanselow, 1992; Anagnostaras et al., 1999).

Over the past several decades, researchers have begun to explore the molecular and cellular bases underlying memory formation. Much of our knowledge stems from the analysis of long-term potentiation (LTP) in the CA1 region (Bliss and Collingridge, 1993; Malenka and Nicoll, 1999). The induction of the major forms of LTP and long-term depression (LTD) requires the activation of the NMDA receptor. Recent experiments provide strong evidence for the role of the NMDA receptor in memory formation. For example, the genetic knockout of NMDA receptor function in the mouse CA1 region leads to severe deficits in both spatial and nonspatial learning (Tsien et al., 1996; Rampon and Tsien, 2000; Huerta et al., 2000). Furthermore, the genetic enhancement of NMDA receptor function, by overexpressing the NR2B subunit in the forebrain, results in superior performance in six different learning and memory tests (Tang et al., 1999, 2001). However, the contribution of the NMDA receptor to various phases of the memory process, such as acquisition, consolidation, and storage, has not been clearly differentiated. In addition, the molecular and cellular mechanisms underlying the formation of declarative long-term memory have not been fully characterized.

Investigation of the consolidation of long-term declarative memory has been approached by two parallel levels: namely, the molecular level and the network/computational level. Because of technical limitations of both approaches, as well focusing on different parts of problems, a common framework underlying memory consolidation has not yet emerged.

At the molecular level, the formation and consolidation of long-term memory are thought to be ultimately expressed in the form of structural changes at synapses. One leading hypothesis is that the formation of long-term memory is the result of a molecular cascade triggered by memory acquisition (Kandel et al., 2000). Such a cascade would consist of receptor activation, transient changes in protein phosphorylation, new protein synthesis, and gene expression, leading to morphological

Grant sponsor: W.M. Keck Foundation; Grant sponsor: Burroughs Wellcome Fund; Grant sponsor: ECNU Foundation; Grant sponsor: National Institute of Mental Health.
*Correspondence to: Joe Z. Tsien, Department of Molecular Biology, Princeton University, Princeton, NJ 08544. E-mail: jtsien@princeton.edu
Accepted for publication 10 May 2002
DOI 10.1002/hipo.10102

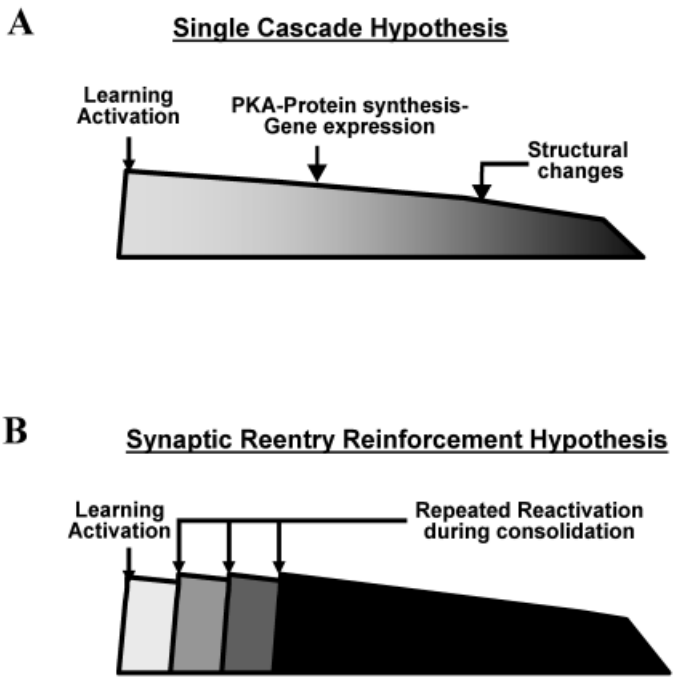

FIGURE 1. A: Traditional view of memory formation represented by the single cascade hypothesis. Learning triggers receptor activation followed by the activation of various kinases and phosphatases, protein synthesis, and gene expression. This molecular cascade results in structural changes underlying long-term memory. B: Here we propose the synaptic reentry reinforcement (SRR) hypothesis. Instead of a single cascade, repeated reinitiation of the NMDA receptor-gated cascade is necessary for the consolidation and storage of long-term memory in the mammalian brain.

changes of neural circuitry (Fig. 1A). However, the time scale of a single LTP-like molecular cascade is not adequate to account for long-term memory consolidation, known to continue for weeks, months, or even years after the initial learning experience. Moreover, individual synaptic receptor proteins are known to be degraded over the course of days in vivo (Shimizu et al., 2000). It is unclear how changes resulting from a single molecular cascade can be sustained in the face of such dynamic turnover of synaptic receptors and their associated structural and signaling proteins.

At the network level, a common approach to understanding the hippocampal memory system has been the application of computational methods (Tsodyks et al., 1996; Shen and McNaughton, 1996; Hasselmo and McClelland, 1999; Lisman and Otmakhova, 2001). The concept of memory consolidation in this context generally refers to the conversion of newly formed declarative memories to cortical memories independent of the hippocampus. Several models have been proposed to explore the relationship between the hippocampus and cortex during the consolidation of memories (Alvarez and Squire, 1994; McClelland and Goddard, 1996; Shen and McNaughton, 1996; Kali and Dayan, 2000). Many of these models propose that the reactivation of hippocampal activity drives the strengthening of synaptic connections in cortical areas. Such work captures the essence of experimental evidence for the critical role of the hippocampus in long-term memory formation (Zola-Morgan et al., 1986; Kim and Fanselow, 1992; Wilson and McNaughton, 1994; Anagnostaras et al., 1999; Riedel et al., 1999; Shimizu et al., 2000).

Recent experiments using an inducible and CA1-specific gene knockout technique showed that long-term memory consolidation requires continued NMDA receptor function in CA1 during the initial weeks after training (Shimizu et al., 2000) (Fig. 2). Because the CA1 NMDA receptor is a major molecular switch for synaptic plasticity, (Bliss and Collingridge, 1993; Bear and Malenka, 1994; Nicoll and Malenka, 1999), we have concluded that the NMDA receptor needs to be reactivated during memory consolidation, thus triggering additional rounds of synaptic modification. This process, termed synaptic reentry reinforcement (SRR), allows further strengthening of neuronal connections during the consolidation period. Such off-line reactivation of the NMDA receptor, both in the hippocampus and cortex, may provide a general cellular mechanism for the consolidation of long-term memory. Furthermore, the occurrence of SRR in cortex could also provide a mechanism for the long-term maintenance of old memories in the mammalian brain.

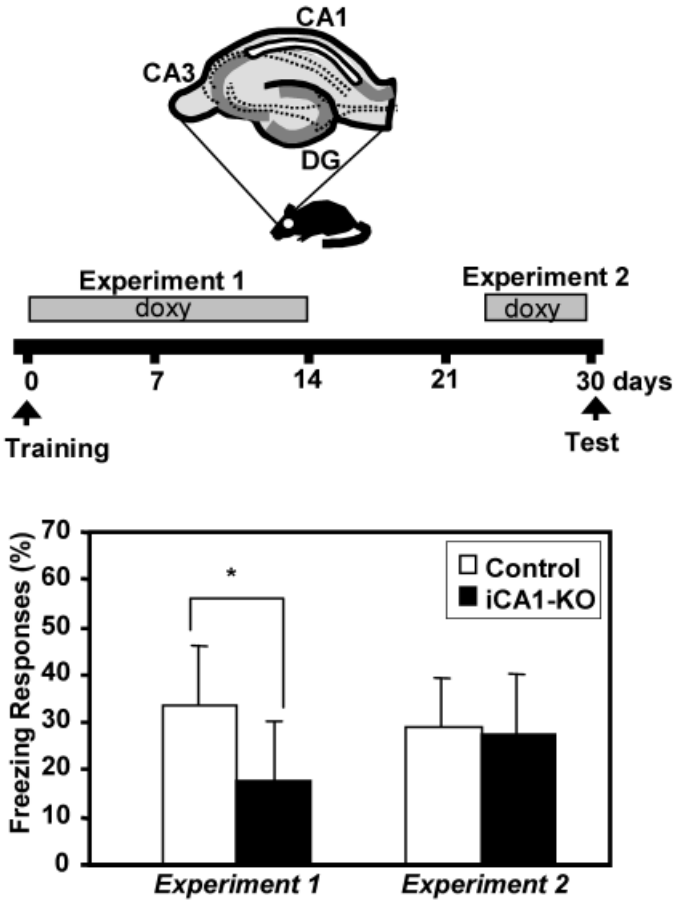

FIGURE 2. Inducible knockout of the CA1 NMDA receptor in mouse hippocampus in the initial weeks post-training leads an impairment in 1-month-old hippocampal memories (contextual fear memory shown here in experiment 1). Whereas knockout of the CA1 NMDA receptor at late stage, before retrieval does not impair memory recall (experiment 2). (Data from Shimizu et al., 2000.)

In light of the potential role for SRR as a crucial process during memory consolidation, the present article explores the consequences of continued hippocampal plasticity during the reactivation of memories. This allows us to demonstrate the effect of multiple rounds of NMDA receptor reactivation, as opposed to treating the hippocampus as a locus for one-shot learning. First, we describe the basic properties of our computational model and examine the consequences of SRR on a single memory stored in an associative memory network. Next, we consider the case where we store multiple memories in the network, showing how the reinforcement of memories during reactivation within the hippocampus could play a role in the further processing of memories. Finally, we examine the effect of SRR on a hippocampal-cortical model for memory consolidation in cortex. We believe that such a computational model, incorporating both molecular and network features, may help us build a general framework describing the role of the hippocampus in long-term memory.

SRR NETWORK MODEL

Network Dynamics and Architecture

We have modeled the process of memory consolidation by incorporating features of the SRR process into the classical attractor memory model (Cohen and Grossberg, 1983; Hopfield, 1984). We built our model to capture two key features revealed by the experiments of Shimizu et al. (2000). First, we incorporate NMDA receptor-dependent plasticity extending beyond initial memory acquisition. To model the involvement of the NMDA receptor in synaptic modification, we include the dynamics of the Hebbian learning rule, $\frac{dw_{ij}}{dt} \propto \eta V_i V_j$ in which correlated firing between pre- and postsynaptic neurons leads to NMDA receptor-mediated changes in synaptic strength. Second, we interpret intrinsic processes such as the dynamic turnover of synaptic proteins (Shimizu et al., 2000) as having a destabilizing effect on stored memory traces. To account for such processes, we include a global synaptic decay term of the form $\frac{dw_{ij}}{dt} \propto -\gamma w_{ij}$.

In all simulations performed, the network evolves according to the following system of differential equations:

$$\tau_u \frac{du_i}{dt} = -u_i + \sum_j w_{ij} \tanh(\beta u_j) + I_i \qquad (1)$$

$$\Delta w_{ij} = -\bar{\gamma} w_{ij} + \bar{\eta} V_i V_j \quad \text{or} \quad \tau_w \frac{dw_{ij}}{dt} = -\gamma w_{ij} + \eta V_i V_j \qquad (2)$$

where u_i represents the membrane potential, and $V_i = \tanh(\beta u_i)$, the firing rate of neuron i. This is a simplification of true neural dynamics in which V_i can be thought of as a moving time average of the neuron's instantaneous firing rate, where averaging is performed over an interval of the same order as the membrane time constant τ_u. Furthermore, the firing rate, V_i, can be expressed as deviations from the basal firing rate, taking on both positive and negative values.

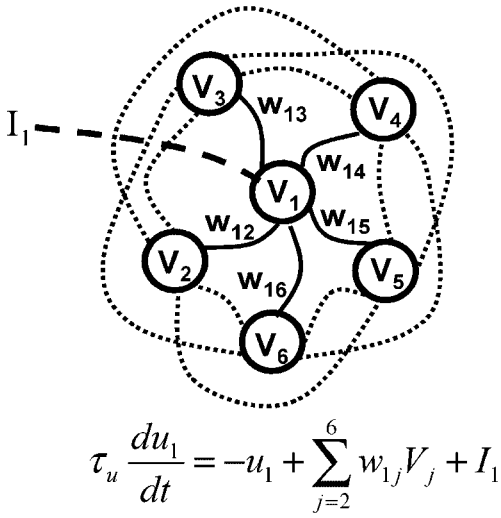

FIGURE 3. Network variables in a small fully connected network. The equation governing the update of the membrane potential of neuron 1 is also shown. The first term represents the decay of the membrane potential, $-u_1$ toward its resting value. The second term is a weighted sum of firing rates, $V_j = \tanh(\beta u_j)$, from neurons presynaptically connected to neuron 1, where the strength of the connection between two neurons is denoted as w_{ij}. The final term I_i represents external inputs to neuron 1, either from other network layers or sensory inputs during learning.

The first term in equation (1), $-u_i$, causes the membrane potential to decay exponentially to its resting potential (which we take to be 0 mV, for simplicity) in the absence of synaptic inputs, just as we expect for true biological neurons. The second term, $\sum_j w_{ij} \tanh(\beta u_j)$, represents changes in the membrane potential of a given neuron, i, due to the firing rate of each presynaptic neuron, j, which we are modeling explicitly, weighted by the strength of the synapse connecting them, w_{ij}. The final term, I_i, corresponds to synaptic inputs coming in from other network layers for which neurons are not explicitly modeled. For instance, in our model of the hippocampal network, the second term represents inputs from other neurons within the hippocampal layer, while I_i denotes the net input to neuron i due to cortical afferents. Network variables are illustrated in Figure 3.

In this model, we allow w_{ij} to take on both positive and negative values. Although in the hippocampus, the true connection between a pair of pyramidal neurons can never be less than zero, it is important to note that we are modeling the effective connection strength between each pair of neurons. Thus, in reality a negative connection would represent a more complex interaction between excitatory neurons and interneurons, which, for simplicity, we do not model explicitly. Instead, we consider a single population of neurons that are capable of forming both excitatory and inhibitory connections. Furthermore, requiring η, γ, and τ_w to take on the same value for all neurons results in the symmetry of synaptic weights such that given any two neurons, i and j, $w_{ij} = w_{ji}$ provided these variables were initialized symmetrically.

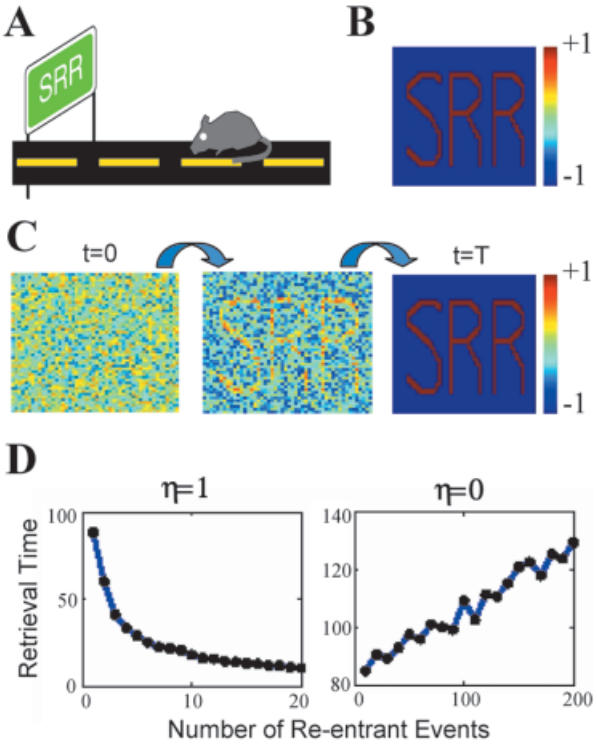

FIGURE 4

In the experiments described below, the weight matrix will be updated either discretely, after an attractor state has been reached, or continuously, as the network evolves. All networks modeled are fully connected within a given layer representing either a well-connected group of hippocampal or cortical neurons. Where we model a two-layer, hippocampal-cortical network, we use one-to-one mapping for the connections between the two layers neglect-

FIGURE 4. Synaptic reentry reinforcement (SRR) process provides a consolidation mechanism for both the strengthening and maintenance of memories. A: Memory acquisition occurs during awake behavior, providing sensory input, which drives synaptic change in the brain. B: In our network model, a learned binary pattern, I (with $I_i \in [-1,+1]$), forming the letters SRR, is stored in the network by setting initial weights according to $w = (0.004)I^T I$. C: During recall, neural activities are randomly initialized, and the network evolves until the memory trace, the letters SRR, is retrieved. The strength of the memory trace is represented by the amount of time for the network to settle to this attractor state. D: Time to reach the memory from a random initialization of the network decreases exponentially as a function of SRR events (left). Whereas in the absence of SRR ($\eta = 0$), in time, the memory decays away as indicated by an increasing long amount of time to settle to that attractor state (right). Data shown are the average of 10 simulations.

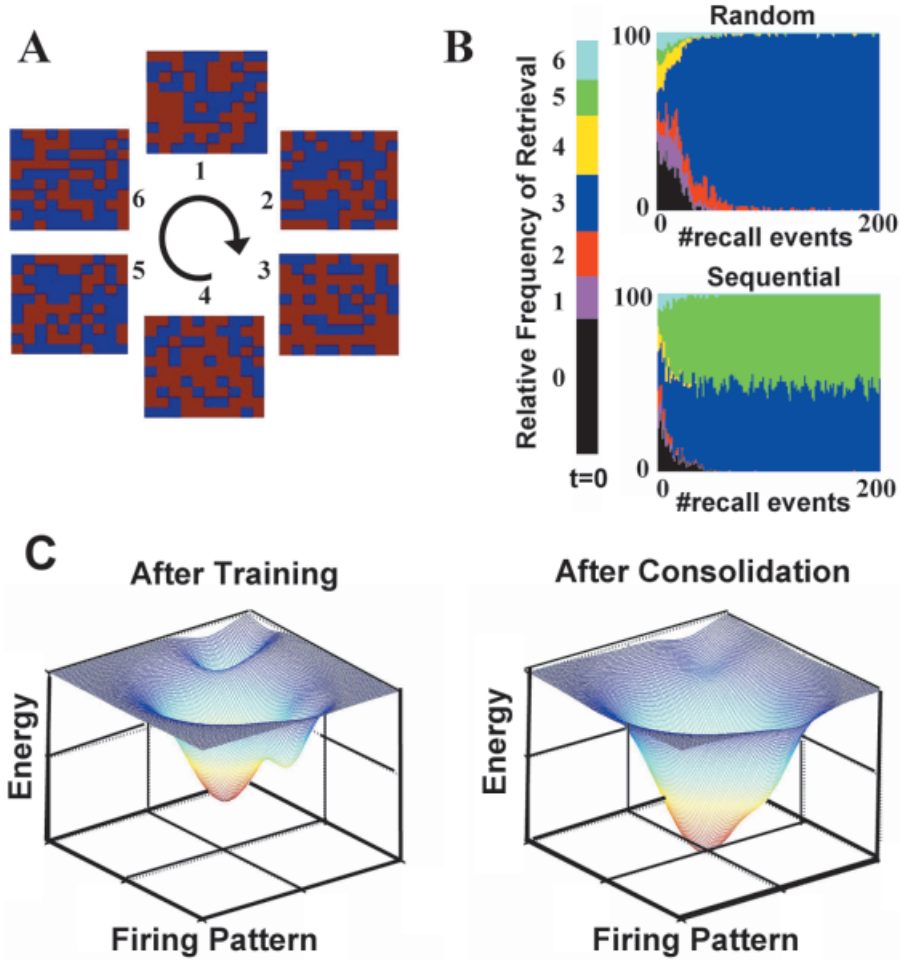

FIGURE 5

ing at this point any feed forward processing, as well as any learning at this synapse, both of which we would expect to find in the biological system.

The dynamics of equations (1) and (2) can be described by an energy function, E, the existence of which shows us that in the presence of stationary inputs, both the dynamics of the neuronal firing rates, and the values of synaptic weights will converge to a stable state. The energy function for the neural dynamics alone, in the presence of constant weights, is obtained by dropping the bracketed term:

$$E = -\frac{1}{2}\sum_{ij} w_{ij}V_iV_j + \sum_i \int u_i dV_i - \sum_i I_iV_i + \left(\frac{1}{4}\frac{\gamma}{\eta}\sum_{ij} w_{ij}^2\right) \quad (3)$$

In such an attractor model, a memory trace is registered by stable activity patterns of the neurons in the network. These patterns lie at the fixed points of the dynamical system (Eq. 1) that is, $\frac{d\mathbf{u}}{dt} = 0$ which are also the local minima of the energy function (Eq. 3). The breadth of an attractor represents the space of initial conditions, that is the initial values of $u_i(t = 0)$, which will lead to the eventual stable activation of a particular memory trace. In our model, the depth of the attractor is directly related to the speed with which a memory trace will be reached.

FIGURE 5. Effect of synaptic reentry reinforcement (SRR) on the consolidation of multiple memory traces in a recurrently connected network. A: During training, binary patterns ($I_i = -80$ or $+80$) are presented to the network, in a sequential manner, clockwise from one to six. B: Selective consolidation of memory traces based on reactivation strategy. Left bar: Vertical height of each color in the left multicolor bar represents percentage of initial conditions leading to the eventual reactivation of that memory trace (or another "junk" attractor shown in black, 0) immediately after training. This gives an estimate of the relative sizes of the attraction basins of each memory trace created by memory acquisition. In addition, this illustrates the $t = 0$ point of the right panels, which show how the relative strengths of the memory traces evolve in time as the network undergoes consolidation driven by two different reactivation schemes. Upper right: The network is periodically reactivated at random and allowed to settle into one of the memory states. This results a competition between patterns, in which each memory strength is represented by its chance to settle in an attractor state (each color occupies certain percentage between zero and 100%, y-axis) and over time of consolidation (x-axis) only one memory, exemplified by memory 3 in blue, is eventually stored in the network (with blue pattern reaching 100% successful retrieval). Lower right: Rather than random reactivation, the network is initiated alternately near two attractor states exemplified by memories 3 and 5, represented in blue and green, respectively. As a result, both attractors remain stable. C: Illustration of the effect of synaptic reentry reinforcement on the energy function during the consolidation process. The landscape is shaped by the synaptic efficacies, w_{ij}. Each point on the x,y-axis represents a specific firing pattern of the neurons. The local minima correspond to attractor memory states. Left: Illustration of an energy function created by learning, representative of the memories stored in the synaptic efficacies immediately after acquisition. In this example we show three memory traces stored as local minima of the energy function. Right: As consolidation proceeds with synaptic reentry reinforcement, major memories are preferable strengthened. Their attractors become deeper and broader, at the expense of the stability of other more weakly stored memories.

We consider two stages of hippocampal memory formation: initial memory acquisition, and memory trace reactivation. To model memory acquisition, a set of inputs are presented to the network to mimic the presence of sensory inputs while an animal is learning a new task (Fig. 4A). After the training period has ceased, and there are no external inputs to the system, $I_i = 0$, the network dynamics remain unchanged (Eqs. 1 and 2). From this stage, our model deviates from common practice in other published work in that reactivation of the network continues to activate the NMDA receptor, resulting in the ongoing modification of synaptic efficacies. Using this general architecture of the model, we explore the effect of the SRR process on associative memory by examining the network storage of a single memory trace, multiple memory traces, and finally in the context of memory consolidation between the hippocampus and cortex.

Single Memory Trace in the SRR Network

We begin by demonstrating the effect of repeated rounds of synaptic modification on the storage of a single memory trace in a recurrent network consisting of 2,500 neurons. During the training period, binary inputs ($I_i = -1$ *or* $+1$) forming the letters SRR (Fig. 4B) are presented, and the network evolves according to Equation 1 until reaching steady state. There are two possible methods for adjusting synaptic weights. First, the weight updates can be discrete, occurring only once a memory trace, or stable attractor of the network, has been reached. Second, the weight updates can occur continuously as the firing rates of neurons are evolving in time. In this section, we demonstrate that multiple rounds of synaptic modification favorable to a single memory trace are capable of strengthening or maintaining that memory. For simplicity, we choose the discretized update method, such that each time an attractor state is reached, synaptic weights are incremented by $\Delta w_{ij} = -\bar{\gamma} w_{ij} + \bar{\eta} V_i V_j$, which we define as a single SRR event. Provided $\tau_u << \tau_w$, that is, synapses evolve on a much slower time scale than neuronal firing rates, the two methods are essentially the same.

After each SRR event, the network is randomly initialized, with u_i distributed between -0.5 and $+0.5$, in the absence of external inputs, resulting in the recovery of the single stored pattern (Fig. 4C). Model parameters used in the simulation were as follows: $\tau_u = 1, \beta = 1, \bar{\gamma} = 0.002, \bar{\eta} = \bar{\gamma}, dt = 0.01$. We initialize the network with the memory trace stored weakly as $w_{ij} = (0.004) I_i I_j$, representing the storage of the memory trace after initial memory acquisition, and equivalent to two training iterations of the network under the discretized update method. In the presence of repeated NMDA receptor-mediated synaptic modification, reactivation of a memory trace is sufficient to enhance its strength and stability. Retrieval time, the time from a random initialization of the network, until the memory is reached, provides a measure of the strength of a memory. We show that retrieval time decreases with the number of SRR events (Fig. 4D), until reaching steady state. In terms of the energy function, this corresponds to a deepening of the single attractor in the network. The synaptic weights continue to evolve until reaching a fixed point of the weight dynamics. From examination of the dynamical system (Eqs. 1 and 2),

we see that for high gain neurons this occurs when all w_{ij} have reached a saturating value at $w_{ij} \approx \pm \frac{\eta}{\gamma}$. At this point, NMDA-dependent synaptic reinforcement serves to maintain the representation of the memory trace within the network, countering the deleterious effects of the included decay processes. It is worth pointing out that immediately after training, we do not know how strongly memories are stored in the network. Although it is likely this is something that varies from memory trace to memory trace. As a result, we cannot distinguish between the requirement for repeated NMDA receptor activation as a need to further strengthen learned memory traces or to maintain their representation within the network long enough for memory consolidation to occur.

To simulate the effect of an inducible NMDA receptor knockout, we set η to 0 in Equation 2. In the absence of synaptic reinforcement after initial training, as when the NMDA receptor is switched off during the consolidation period, the newly formed memory trace takes an increasingly long time to be retrieved (Fig. 4D), due to the synaptic decay term, governed by the parameter γ. This demonstrates the necessity of SRR to overcome the synaptic destabilization of memory traces during the consolidation process.

Multiple Memory Traces in the SRR Network

Next, we examine the effect of the SRR process on an associative memory network storing multiple memory traces. We consider this network to represent a small percentage of connected hippocampal neurons, capable of storing multiple memory traces during memory acquisition. We model explicitly both the memory acquisition and reactivation processes. Network parameters for the following simulations are as follows: $\eta = 1$, $\tau_u = 1$, $\tau_w = 1000$, $\beta = 1$, and $\gamma = 1$. To step forward in time, we perform Euler integration of equations (1) and (2) with $dt = 1$. Before initial learning, all synaptic connections within the hippocampal network take on the neutral value, 0, and the membrane potential, u_i, is randomly distributed between ± 1.

During memory acquisition, the synaptic connections of the network are capable of converging to a representation in which multiple memory traces are stored. To demonstrate this, we examine a network of 100 neurons, presented with six random binary input patterns in a sequential, and cyclic manner (Fig. 5A). We regard this as an analogy to an animals experience running down a linear track many times during a given training session, observing cues at various locations en route. We present the network with strong inputs, $I_i \in [-80,80]$, in order to dominate the recurrent connections within the network, driving the network toward an attractor state representing the presented input (Dong and Hopfield, 1992). Each pattern is presented for 12 time steps before switching to the next pattern in the sequence, and this process is continued for 3,000 iterations of the network.

In time the synaptic efficacies of the network, w_{ij}, reach steady state, in which all six memories are stored as fixed points of the dynamical system. That is, initializing the network in any one of the six stored memory states, the network will remain in that state indefinitely. We can estimate the relative size of the basins of attraction for each memory by randomly initializing the network 100 times, and recording how frequently any given memory is retrieved, as shown in Figure 5B (left).

After this initial training period, inputs are switched off, but the learning dynamics continues as the network undergoes a periodic, random reactivation every 12 time steps. After each such reactivation, the network settles into one of the attractor states of the network. In the presence of continued Hebbian learning this memory trace will be strengthened. As a result, the basin of attraction is increased, and on future random initializations of the network this memory will be more likely recovered. In this manner, when reactivation of the network occurs randomly, previously stored memories compete with one another until only a single pattern is strongly stored in the network. This is demonstrated in the top right corner panel of Figure 5B, where after each reactivation step, synaptic weights are held fixed and the relative sizes of attraction basins are measured as described above.

While randomly reactivating the network results in only a single memory being stored, it is not clear that this is the same method used by the brain to reactivate memory traces. First, the hippocampus is a locus that stores sequences of events; therefore, activation of one memory trace should bias the network to activate another particular memory subsequently. Second, recent evidence has shown that correlated pre- and postsynaptic spiking can lead to increased excitability of the presynaptic neuron (Ganguly et al., 2000). This would bias the network to reactivate neurons that have recently been involved in learning. We next choose our periodic reinitializations of the network to be alternately near (within the basin of attraction of) one of two memories stored in the network, with u_i initialized as ± 1. In this case, both memories remain stored in the network, while the others eventually lose stability (Fig. 5B, bottom).

It is worth pointing out that in a network storing multiple memories, an important distinction must be made between Hebbian learning which occurs during training, and that which occurs upon subsequent reactivation. In the former case, inputs to the network can dominate the dynamics of the system, whereas in the latter, it is the synaptic weights that dominate (Dong and Hopfield, 1992). As a result, the SRR process can lead to additional processing of hippocampal memories in a manner that is highly dependent on the way in which memories are reactivated. When one considers the hippocampal dependent consolidation of cortical memories, one might expect that any processing which occurs at the level of the hippocampus during consolidation would have direct effect on which memories will be consolidated at the cortical level for long-term storage.

Hippocampal-Cortical Consolidation in SRR Model

It has been shown that the requirement of hippocampal NMDA receptor reactivation for cortical consolidation is time dependent (Shimizu et al., 2000). To demonstrate this effect, we extend our model to one with an architecture hierarchically similar to that used by Alvarez and Squire (1994). We add to the network described above two "cortical" areas (Fig. 6A), which evolve accord-

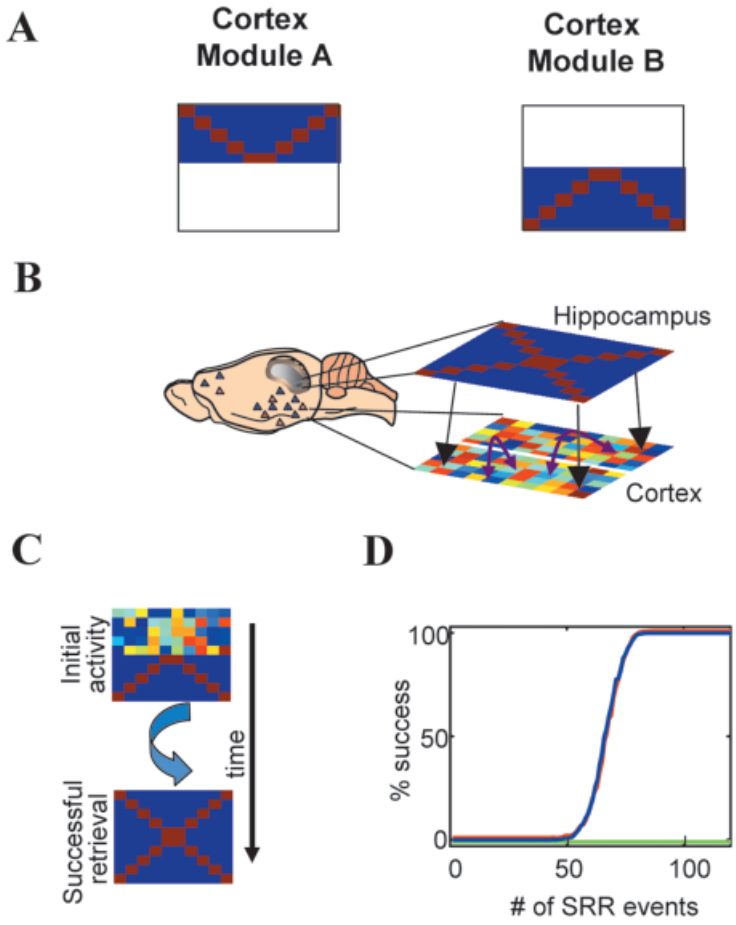

FIGURE 6. Synaptic reentry reinforcement (SRR) within the hippocampus is required in order to provide coordinated output to drive cortical memory consolidation. A: During learning, cortical modules A and B are activated and provide input driving hippocampal neurons. B: During consolidation, the hippocampus (in gray banana-shaped region) reactivates and further strengthens the stored memory trace by SRR. Coherent hippocampal reactivation provides coordinated reactivation of cortical modules, resulting in the SRR-based strengthening of synaptic efficacies primarily between cortical modules A and B, as well as within each module. Colored triangles represent cortical neurons undergoing reactivation and consolidation. C: Cortical consolidation is measured by the time taken to retrieve the memory trace in module A, triggered by the reactivation of the memory trace in module B, or vice versa. The frequency with which the cortical network recovers the full memory is recorded. D: Blue curve demonstrates that the ability of the cortical network to retrieve the full pattern improves with each SRR event. Green curve shows that turning off the SRR process in the hippocampus after initial memory acquisition results in the inability to consolidate the memory in cortex. Red curve shows that turning off the SRR process within the hippocampus at a later stage ($t = 50$) has no effect on the ability of cortex to retrieve the already consolidated cortical memory traces.

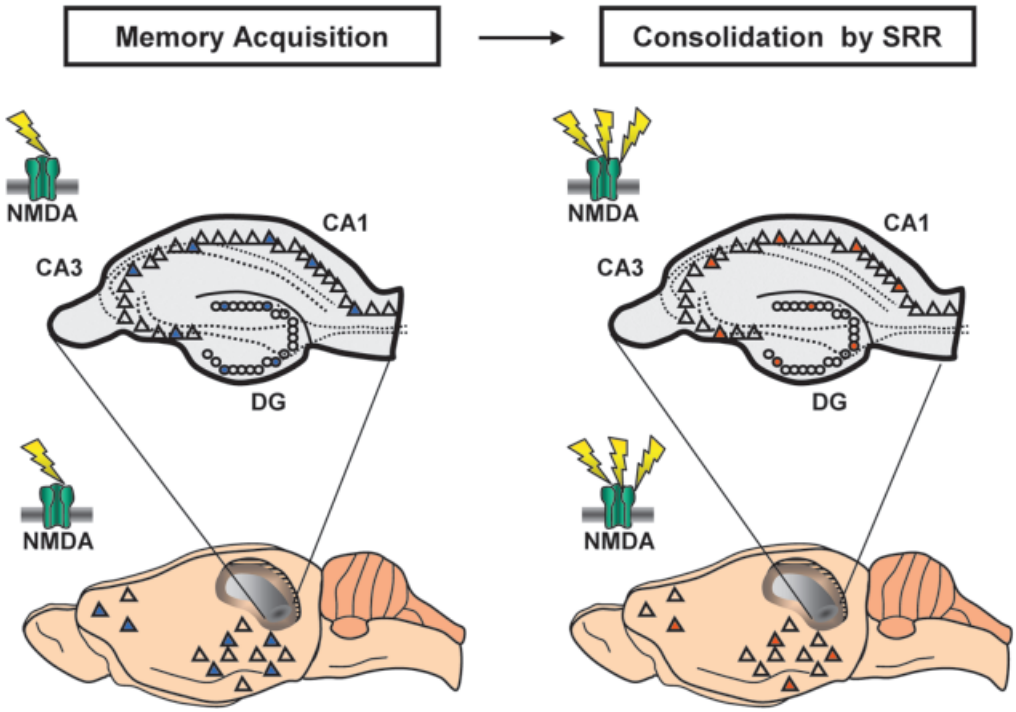

FIGURE 7. Essential role of NMDA receptor activation in both memory acquisition and consolidation. Cartoon (left) illustrates that NMDA receptor activation in a group of neurons (blue) during learning creates memory traces in both the cortex and hippocampus (shaded gray). Cartoon (right) illustrates that multiple reactivations of both hippocampal and cortical NMDA receptors on a set of neurons (red) led to the eventual consolidation of long-lasting memory traces in cortex.

ing to the same basic dynamics described in Equations 1 and 2. For simplicity, we use a one-to-one mapping between hippocampal and cortical neurons, with a connection strength of 80, as we seek to demonstrate only the strengthening of memory traces in cortex, as supported by the hippocampus, without any pretext about information processing which may occur in the feedforward network (Lisman and Otmakhova. 2001; McClelland and Goddard, 1996). Thus, our hippocampal network contains 100 neurons, and each cortical module contains 50 neurons.

Initial synaptic weights within the cortical areas are symmetric, and distributed between -1 and 1, with five patterns initially stored in the network, representing previously stored memory traces. During initial training, input reaches the hippocampus through the cortical network, which receives direct binary inputs $I_i \in [-80,80]$ for 100 iterations. Parameter values are the same as those listed in the previous section with the exception of dt which equals 0.2. After initial memory acquisition, the cortex is tested for its ability to retrieve this memory in the absence of hippocampal input in the following manner. One cortical module is activated in the stored memory state, while the other is activated at random. This is repeated 100 times, and the frequency with which the complete pattern is recovered is recorded.

The effect of SRR is demonstrated by randomly reactivating the hippocampal network every 40 time steps. As the network evolves to the stored memory, it results in the additional reactivation of the cortical memory trace (Fig. 6B). After each such SRR event, we test the performance of the cortical network as described above (Fig. 6C). Figure 6D shows the performance of the cortical network alone as a function of the number of SRR events (blue curve). Post-training inactivation of the hippocampal NMDA receptor ($\eta = 0$) hampers the eventual strengthening of memory traces in cortex (green curve). In contrast, at a later stage, the loss of NMDA receptor function in the hippocampus has little effect (red curve), as cortical memory traces have been sufficiently consolidated (Fig. 6D).

DISCUSSION

Our computational model illustrates the effect of SRR in the process of hippocampal based memory consolidation. We demonstrate that the SRR process can strengthen and maintain memory traces despite inherent variability in the system due to such processes as the turnover of synaptic receptors and structural proteins. The strengthening of a single memory stored in the network is observed as a decrease in the average time to reach the attractor state. Because of the global synaptic decay term in our model $\frac{dw_{ij}}{dt} \propto -\gamma w_{ij}$, only in the presence of sufficient and repeated synaptic reinforcement can memories be maintained and strengthened in the hippocampus. Likewise, multiple rounds of synaptic reinforcement, triggered by repetitive training events, can achieve a similar effect on the strengthening of memory traces in the network.

Next, we considered the alternative scenario in which multiple memory traces are created in a network during training. We modeled this by presenting six memory traces, one at a time, to the network, analogous to several visual cues that an animal might notice while exploring a novel environment. Such a training paradigm results in the formation of a memory attractor for each of the six presented cues. During the reactivation period, in the presence of the SRR process, changes in relative memory strength are shown by examining the relative size of each individual memory trace's attraction basin. For an individual memory trace to be consolidated, SRR must overcome not only the synaptic destabilization processes described, but it must also compete with the strengthening of the other five memory traces. In this case, the consolidation of multiple memory traces is a biased selection process. Here we have considered several theoretical factors that can influence which memory traces will be stable.

First, input biases may develop based on the relative importance of cues to the animal. For example, an especially large amount of time spent exploring a particular cue during a single training event would lead to a stronger memory trace for that cue than for others in our network model. In addition, when an otherwise neutral cue becomes associated with reward or danger, a stronger memory trace is created to represent that cue. In our model this could be represented by an increase in the parameter η during the presentation of that pattern.

Second, differences in the frequency with which memory traces are consciously recalled may serve as another factor biasing which memories are consolidated. For example, in our model, if one particular memory trace is actively recalled, and thus strengthened by the SRR process, then during subsequent consolidation, under the random reactivation paradigm presented, this memory trace will be consolidated preferentially. This is consistent with our own experience, in which the more frequently a particular childhood event is recalled, the better and longer it will be remembered into adulthood.

Third, subconscious reactivation during sleep can also influence the memory selection process. In essence, the relative frequencies with which particular memory traces are replayed during sleep or dreaming play an important role in biasing which memories are consolidated. In addition, the mechanism by which memory traces are reactivated strongly influences memory consolidation. For example, in our model, when the network is periodically reactivated in a random manner, eventually only a single memory trace is strongly stored in the network. In contrast, if reactivation occurs in a sequential manner, with one memory trace always leading to the subsequent reactivation of another, both memories will be consolidated (Fig. 5).

It is interesting to note that despite the inability to recall the weakened patterns on random reactivation, for a very long time they remain fixed points of the dynamical system. Although the basin of attraction has become extremely small, if started in exactly the stored state, the state remains stable. Even after losing stability, the ghost of the attractor still causes the dynamics of the system to slow near the formerly stored pattern (Hopfield, 1982). Thus, the network still contains information that this once was a stored pat-

tern, biasing the network to more easily be capable of relearning this pattern than to learn an entirely new pattern.

Increasing evidence suggests a role for sleep in memory consolidation (for review, see Maquet, 2001; Siegel, 2001; Stickgold et al., 2001). Moreover, learning-induced correlations in the firing of hippocampal place cells have been reported to reappear during sleep (Wilson and McNaughton, 1994; Louie and Wilson, 2001). Such a coordinated reactivation of these neurons suggests the existence of a natural condition in which the NMDA receptor, a cellular coincidence detector, can be reactivated, thus reinforcing the synaptic connections between them.

We propose that SRR is a general mechanism for memory consolidation in multiple brain regions, including the hippocampus, the amygdala, and cortex. This notion is supported by recent evidence that NMDA receptor reactivation is required for the consolidation of memory traces involved in cued fear extinction, a hippocampal independent memory task (Santini et al., 2001). Further, a recent study has shown that impaired LTP in the cortex appears to be correlated with late-stage deficits in cortical memory consolidation (Frankland et al., 2001). These suggest a general requirement for repetitive synaptic reinforcement through multiple LTP-like events in many brain regions.

To account for the role of the hippocampus in converting short-term declarative memory in to long-term memory, we recently proposed that during the consolidation period, the hippocampus serves as a coincidence regenerator for the coordinated reactivation of cortical neurons, activating the cortical NMDA receptor, and strengthening intercortical connections through SRR (Shimizu et al., 2000). Indeed, recent observations show that correlations between hippocampal-cortical neurons emerging during learning can be subsequently measured during sleep (Qin et al., 1997). As demonstrated by our model, this could allow cortical neurons previously corresponding to different sensory modules to be reactivated together, leading to the strengthening of the connections between them. Without coordinated input from the hippocampus, it may be difficult for neurons belonging to different sensory modules to reactivate together after initial memory acquisition. This makes the NMDA receptor reactivation between these neuronal connections less likely, thus preventing the consolidation and binding of cortical memories spanning multiple sensory modules.

Both our experimental and modeling work suggest that the SRR process within the hippocampus is necessary for the occurrence of SRR-mediated consolidation of cortical memories (Fig. 7). In the absence of SRR in the hippocampus, memory traces do not remain stable due to the inherent instability derived from the dynamic turnover of receptors and structural proteins at synapses. As a result, it would be difficult for the hippocampus to maintain its ability to provide coherent output to drive the reactivation of cortical neurons.

The requirement for hippocampal SRR in memory consolidation is known to be time dependent. For instance, the knockout of the CA1 NMDA receptor during the initial 1–2 weeks, but not the 4th week, impairs long-term memory formation (Shimizu et al., 2000). Analogously, in our computational model, setting the parameter η to 0 within the hippocampal network, immediately after initial memory acquisition prevents cortical memory consolidation, whereas turning off this parameter at a later time has no detrimental effect (Fig. 6).

It remains unclear whether cortical consolidation is complete by the time memories have become hippocampal independent. We can imagine two possible scenarios. In the first scenario, cortical consolidation is time-locked with SRR in the hippocampus. This means that cortical memory consolidation has reached its final level by the time the hippocampus is no longer required. In the second scenario, cortical memories will be further strengthened in a hippocampal independent manner through the cortical SRR process, until steady state is reached. More experiments are necessary to differentiate between these two possibilities.

It is also unclear what happens to memory traces left behind in the hippocampus after long-term memories have been stored in the cortex. It has been suggested that previously stored memory traces might interfere with the ability to learn new memories (Hasselmo, 1993). It has also been suggested that the hippocampus may have limited storage capacity, as there are only roughly 200,000–300,000 CA3 and 300,000–400,000 CA1 pyramidal cells, and 700,000–1,000,000 granule cells in rodent hippocampus (Traub and Miles, 1991). It is conceivable that the continued accumulation of outdated memory traces in the hippocampus may overload the system gradually over time, eventually disabling hippocampal function in memory consolidation. A typical way this problem has been dealt with in computational models is to clamp the network to the new inputs by either providing very strong inputs, or simply ignoring the effect of previously modified synapses.

Recently, we postulated a cellular mechanism by which ongoing turnover of adult-born neurons (via adult neurogenesis) in the dentate gyrus can represent a powerful mechanism for the clearance of outdated hippocampal memory traces (Feng et al., 2001). Adult-generated neurons within the dentate gyrus are known to form synapses rapidly with CA3 pyramidal neurons (Gaarskjaer, 1986; Hastings and Gould, 1999). It is feasible that such changes in network architecture will alter the attractor states corresponding to memories previously stored in the network, and in time destabilize them. Interestingly, these neurons are short-lived, typically with a life span of 3 weeks in rodents, which correlates well with the duration of hippocampal dependence of declarative memories. Consistent with the "adult neurogenesis-memory clearance" hypothesis, we have shown that deficient dentate neurogenesis in forebrain-specific *Presenilin-1* knockout mice is associated with reduced clearance of 2-week-old contextual fear memories (Feng et al., 2001).

We further postulate that ongoing SRR in the brain plays a key role in ensuring the long-term storage of cortical memories. This is largely based on the consideration cortical synapses will also be subject to destabilizing forces resulting from processes such as the turnover of synaptic receptors. For example, it is known that the turnover time for the NMDA receptor in vivo is approximately 5 days (Shimizu et al., 2000). It is very likely that other synaptic proteins, such as the AMPA receptor, will also be turned over periodically. This raises the question of how synaptic efficacies can be maintained in a manner that preserves the delicate stored memories. As demonstrated in our hippocampal network, such degra-

datory processes can be successfully countered by the periodic NMDA receptor reactivation and strengthening of memories through the SRR process. Without such a process, over time, synaptic efficacies cannot be stably preserved and memory traces become unreliable, undermining long-term storage of information in the brain.

It is conceivable that abnormal SRR process may contribute to the pathological phenomenon of several neurological disorders. For example, in epileptic patients, certain neurons tend to be more easily excitable, leading to pathological reactivation of those circuits, initiating activation of the NMDA receptor and the abnormal strengthening of synapses in those regions. This abnormal strengthening of synapses could explain why, in certain patients, repeated focal seizure events tend to spread to cortical areas adjacent to the initial focal site, resulting in more generalized seizures.

Kindling is an experimental form of epilepsy. A high-frequency stimulus delivered to the limbic system, which is initially a subthreshold for the production of seizures, becomes capable of triggering intense seizure after repetitive subthreshold stimulation over days and weeks (McNamara et al., 1980). Similar to long-term potentiation, the induction of kindling requires NMDA receptor activation. The NMDA receptor mediated SRR process could contribute to the strengthening of synapses, consequently decreasing the stimulus threshold for the generation of seizure.

Abnormal SRR, in theory, may also account for certain aspects of schizophrenia. Recent observations have shown altered NMDA receptor function, in both hippocampus and cortex, in schizophrenic patients. For example, upregulation of NMDA receptor subunits, such as NR2D, has been found in the prefrontal cortex of schizophrenic patients (Akbarian et al., 1996). This is particularly interesting because the NR2D subunit containing NMDA receptor tends to remain open for a longer duration and has little magnesium dependence, facilitating the activation of the NMDA receptor to a higher degree than the normal NMDA receptor at lower membrane potentials. Excessive levels of the NR2D subunit would decrease the coincidence detection ability of the NMDA receptor, leading to aberrant receptor activation. It has been demonstrated through computational modeling that a lower threshold for the activation of the NMDA receptor can lead to erroneous synapse formation during learning in an associative memory network (Greenstein-Messica and Ruppin, 1998). We postulate that in schizophrenic patients, abnormal reactivation of the cortical NMDA receptor during the consolidation process may produce a situation in which unrelated experiences, or even imagined ones, are aberrantly bound together due to hyperactivity of the SRR process. This may explain certain delusional aspects of the syndrome.

In summary, we have presented a computational model that attempts to bridge the gap between the understanding of memory consolidation at the molecular level vs. the network level. Our model describes the SRR-mediated consolidation of memories within the hippocampus, as well as in the hippocampal-cortical, and cortical-cortical circuitries. We have shown that despite intrinsic drift of synaptic efficacy over time in the brain, occurring over the course of week(s) and beyond, SRR is capable of strengthening and maintaining memory traces, where they would otherwise become unstable over time. Furthermore, we have demonstrated that multiple rounds of synaptic modification triggered by memory reactivation, either during conscious recall or sleep, could lead to the selective consolidation of a subset of memory traces at both the hippocampal and cortical level. In addition, our model predicts an essential role for the SRR process in the maintenance and storage of long-lasting cortical memories, as well as the general stability of neural circuitry in the brain.

Acknowledgments

The authors thank John Hopfield and Olly Downs for stimulating discussion and thoughtful input. This research was supported funding from NIMH, Burroughs Welcome Fund, and W.M. Keck Foundation (to J.Z.T.). The MatLab code for the simulations in this article is available at http://www.molbio.princeton.edu/labs/tsien.

REFERENCES

Akbarian S, Sucher NJ, Bradley D, Tafazzoli A, Trinh D, Hetrick WP, Potkin SG, Sandman CA, Bunney WE Jr, Jones EG. 1996. Selective alterations in gene expression for NMDA receptor subunits in prefrontal cortex of schizophrenics. J Neurosci 16:19–30.

Alvarez P, Squire LR. 1994. Memory consolidation and the medial temporal lobe: a simple network model. Proc Natl Acad Sci U S A 91: 7041–7045.

Anagnostaras SG, Maren S, Fanselow MS. 1999. Temporally graded retrograde amnesia of contextual fear after hippocampal damage in rats: within-subjects examination. J Neurosci 19:1106–1114.

Bear MF, Malenka RC. 1994. Synaptic plasticity: LTP and LTD. Curr Opin Neurobiol 4:389–399.

Bliss TVP, Collingridge GL. 1993. A synaptic model of memory—long term potentiation in the hippocampus. Nature 361:31–39.

Cohen M, Grossberg S. 1983. Absolute stability of global pattern-formation and parallel memory storage by competitive neural networks. IEEE Trans Syst Man and Cyber 13:815–826.

Dong DW, Hopfield JJ. 1992. Dynamic properties of neural networks with adapting synapses. Network 3:267–283.

Feng R, Rampon C, Tang YP, Shrom D, Jin J, Kyin M, Sopher B, Martin GM, Kim SH, Langdon RB, Sisodia SS, Tsien JZ. 2001. Deficient neurogenesis in forebrain-specific presenilin-1 knockout mice is associated with reduced clearance of hippocampal memory traces. Neuron 32:911–926.

Frankland PW, O'Brien C, Ohno M, Kirkwood A, Silva AJ. 2001. Alpha-CaMKII-dependent plasticity in the cortex is required for permanent memory. Nature 411:309–313.

Gaarskjaer FB. 1986. The organization and development of the hippocampal mossy fiber system. Brain Res 396:335–357.

Ganguly K, Kiss L, Poo MM. 2000. Enhancement of presynaptic neuronal excitability by correlated presynaptic and postsynaptic spiking. Nat Neurosci 3:1018–1026.

Greenstein-Messica A, Ruppin E. 1998. Synaptic runaway in associative networks and the pathogenesis of schizophrenia. Neural Comput 10: 451–465.

Hasselmo ME. 1993. Acetylcholine and learning in a cortical associative memory. Neural Comput 5:32–44.

Hasselmo ME, McClelland JL. 1999. Neural models of memory. Curr Opin Neurobiol 9:184–188.

Hastings NB, Gould E. 1999. Rapid extension of axons into the CA3 region by adult-generated granule cells. J Comp Neurol 413:146–154.
Hopfield JJ, 1982. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A 79:2554–2558.
Hopfield JJ. 1984. Neurons with graded response have collective computational properties like those of two state neurons. Proc Natl Acad Sci U S A 81:3088–3092
Huerta PT, Sun LD, Wilson MA, Tonegawa S. 2000. Formation of temporal memory requires NMDA receptors within CA1 pyramidal neurons. Neuron 25:473–480.
Kali S, Dayan P. 2000. Hippocampally-dependent consolidation in a hierarchical model of neocortex. NIPS 2000:24–30.
Kandel ER, Schwartz JH, Jessel TM. 2000. Principles of neural science. 4th ed. New York: McGraw-Hill. 1414 p.
Kim JJ, Fanselow MS. 1992. Modality specific retrograde amnesia of fear. Science 256:675–677.
Lisman JE, Otmakhova NA. 2001. Storage, recall, and novelty detection of sequences by the hippocampus: elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. Hippocampus 11:551–568.
Louie K, Wilson MA. 2001. Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. Neuron 29:145–156.
Malenka RC, Nicoll RA. 1999. Long-term potentiation—a decade of progress? Science 285:1870–1874.
Maquet P. 2001. The role of sleep in learning and memory. Science 294:1048–1052.
McClelland JL, Goddard NH. 1996. Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. Hippocampus 6:654–665.
McGaugh JL. 2000. Memory—a century of consolidation. Science 287: 248–251.
McNamara JO, Byrne MC, Dasheiff RM, Fitz JG. 1980. The kindling model of epilepsy: a review. Prog Neurobiol 15:139–159.
Muller GE, Pilzecker A. 1900. Experimentelle Beiträge zur Lehre vom Gedächtnis. Z Psychol Ergänzungsband 1:1–300.
Nicoll RA, Malenka RC. 1999. Expression mechanisms underlying NMDA receptor-dependent long-term potentiation. Ann N Y Acad Sci 868:515–525.
Qin Y-L, McNaughton BL, Skaggs WE, Barnes CA. 1997. Memory reprocessing in corticocortical and hipocampocortical neuronal ensembles. Philos Trans R Soc Lond B 352:1525–1533.
Rampon C, Tsien JZ. 2000. Genetic analysis of learning behavior-induced structural plasticity. Hippocampus 10:605–609.
Rempel-Clower NL, Zola SM, Squire LR, Amaral DG. 1996. Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. J Neurosci 16:5233–5255.
Riedel G, Micheau J, Lam AGM, Roloff EvL, Martin SJ, Bridge H, de Hoz L, Poeschel B, McCulloch J, Morris RGM. 1999. Reversible neural inactivation reveals hippocampal participation in several memory processes. Nat Neurosci 2:898–905.
Santini E, Muller RU, Quirk GJ. 2001. Consolidation of extinction learning involves transfer from NMDA-independent to NMDA-dependent memory. J Neurosci 21:9009–9017.
Scoville WB, Milner B. 1957. Loss of recent memory after bilateral hippocampal lesions. J Neurol Neurosurg Psychiatry 20:11–21.
Shen B, McNaughton BL, 1996. Modeling the spontaneous reactivation of experience-specific hippocampal cell assemblies during sleep. Hippocampus 6:685–692.
Shimizu E, Tang Y-P, Rampon C, Tsien JZ. 2000. NMDA receptor-dependent synaptic reinforcement as a crucial process for memory consolidation. Science 290:1170–1174.
Siegel JM. 2001. The REM sleep-memory consolidation hypothesis. Science 294:1058–1063.
Stickgold R, Hobson JA, Fosse R, Fosse M. 2001. Sleep, learning, and dreams: off-line memory reprocessing. Science 294:1052–1057.
Sutherland GR, McNaughton B. 2000. Memory trace reactivation in hippocampal and neocortical neuronal ensembles. Curr Opin Neurobiol 10:180–186.
Tang YP, Shimizu E, Dube GR, Rampon C, Kerchner GA, Zhuo M, Liu G, Tsien JZ. 1999. Genetic enhancement of learning and memory in mice. Nature 401:63–69.
Tang YP, Wang H, Feng R, Kyin M, Tsien JZ. 2001. Differential effects of enrichment on learning and memory function in NR2B transgenic mice. Neuropharmacology 41:779–790.
Traub RD, Miles R. 1991. Neural networks of the hippocampus. New York: Cambridge University Press. 4 p.
Tsien JZ. 2000. Linking Hebb's coincidence-detection to memory formation. Curr Opin Neurobiol 10:266–273.
Tsien JZ, Huerta PT, Tonegawa S. 1996. The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. Cell 87:1327–1338.
Tsodyks MV, Skaggs WE, Sejnowski TJ, McNaughton BL. 1996. Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. Hippocampus 6:271–280.
Wilson MA, McNaughton BL. 1994. Reactivation of hippocampal ensemble memories during sleep. Science 265:676–679.
Zola-Morgan S, Squire LR, Amaral DG. 1986. Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. J Neurosci 6:2950–2967.