The site of vesicle fusion, producing a hotspot of channel activation, produces the highly stereotyped AMPA receptor current observed at CA1 synapses. However, the variability of the quantal responses has posed a challenge for computational models of synaptic transmission, which have previously relied on a smaller range of AMPA receptor density at CA1 synapses. Using this value, our simulations correctly predict a quantal amplitude of ~10 pA. We have also analyzed the properties of excitatory postsynaptic currents (EPSCs) generated by the multivesicular release that can occur during evoked responses. We find that summation is nearly linear and that the existence of multiple narrow peaks in amplitude histograms can be accounted for. It has been unclear how to reconcile the existence of these narrow peaks, which indicate that the variation of quantal amplitude is small (CV < 0.2) with the highly variable amplitude of miniature EPSCs (mEPSCs; CV ~ 0.6). According to one theory, mEPSC variability arises from variation in vesicle glutamate content. However, both our modeling results and recent experimental results indicate that this view cannot account for the observed rise time/amplitude correlation of mEPSCs. In contrast, this correlation and the high mEPSC variability can be accounted for if some mEPSCs are generated by two or more vesicles released with small temporal jitter. We conclude that a broad range of results can be accounted for by simple principles: quantal amplitude (~10 pA) is stereotyped, some mEPSCs are multivesicular at moderate and large synapses, and evoked responses are generated by quasi-linear summation of multiple quanta.

**INTRODUCTION**

Rapid progress has been made in understanding the generation of fast excitatory postsynaptic currents (EPSCs) at central glutamatergic synapses (Clements 1996; Dingledine et al. 1999; Edmonds et al. 1995; Jonas and Spruston 1994). Much of the experimental work on transmission and plasticity at central synapses has focused on the Schaeffer collateral synapses of the CA1 hippocampal region, but important properties of quantal transmission at these synapses remain unclear.

One area of uncertainty is the variability of the quantal response. At the neuromuscular junction, quantal responses have stereotyped amplitude, as quantified by the coefficient of variation (CV = SD/mean), which is <0.2 (Boyd and Martin 1956). It has been suggested that quantal responses at hippocampal synapses are much more variable (CV ~ 0.6) (Bekkers et al. 1990; Forti et al. 1997; Liu and Tsien 1995; McAllister and Stevens 2000) or at multiple synapses electrotonically close to the recording electrode (Magee and Cook 2000). It is unclear how this high variability can be reconciled with the observations of narrow peaks in the histograms of evoked responses at CA1 synapses (Bolshakov et al. 1997; Foster et al. 1991; Kullmann and Nicoll 1992; Larkman et al. 1991; Liao et al. 1992; Malinow 1991; Stricker et al. 1996), which indicates that the CV is low (~0.2) and that quantal size is in the ~10-pA range (when measured close to the site of generation (Smith et al. 2003)).

A second question concerns the interaction of multiple quanta. Work on several types of synapses has demonstrated that an action potential can release multiple vesicles at the same active zone (Korn et al. 1993; Prange and Murphy 1999; Wadiche and Jahr 2001; Wall and Usowicz 1998), and similar findings have been made at CA1 synapses (Conti and Lisman 2003; Oertner et al. 2002; Tong and Jahr 1994a). The AMPA current during an average miniature EPSC (mEPSC) is far from saturation (100–200 pA) (Isikawa et al. 2002; Liu et al. 1999; McAllister and Stevens 2000), raising the question of how the responses to multiple quanta summate at individual synapses.

To understand quantal transmission, it is necessary to determine the spatiotemporal pattern of channel opening, which can only be determined by computer simulation (Barbour 2001; Bartol et al. 1991; Faber et al. 1992; Franks et al. 2002, 2003; Holmes 1995; Trommershauser et al. 1999; Uteshev and Penefather 1996; Wahl et al. 1996). New experimental results provide additional constraints on such computational models. First, dendritic recordings provide improved measurements of the synaptic response itself (Magee and Cook 2000) and reveal that the 20–80% rise time is 50–100 µs, much faster than previously thought. Second, a more accurate model of AMPA receptor activation can be constructed using new information about channel properties (Abele et al. 2000; Andrasfalvy and Magee 2001; Armstrong and Gouaux 2000; Robert and Howe 2003; Rosenmund et al. 1998; Smith and Howe 2000). Finally, recent results (Matsuzaki et al. 2001) make it possible to...
estimate the density of AMPA channels on CA1 mushroom spines, a critical determinant of response amplitude. We have developed Monte Carlo simulations of vesicular release of glutamate into the cleft and the consequent activation of AMPA channels. This model correctly predicts the rise time, amplitude, and variance (as deduced from histograms of the evoked response) of quantal responses. Simulations were then used to analyze experimental observations on evoked and spontaneous events. Of particular importance is the analysis of the rise-time amplitude correlation of mEPSCs, a finding not previously explored in models. Our results suggest a simple set of rules for quantal transmission that provides a unified framework for understanding spontaneous and evoked responses.

METHODS

Kinetic scheme

Early models of AMPA receptor kinetics assumed that the channel opened to a single conducting state on binding of two ligand molecules (Jonas et al. 1993). Since then, much more has been learned about the structure, activation, and desensitization of AMPA channels. The channels contain four subunits, each of which has a binding site for glutamate. Structural studies indicate that agonist binding causes a closure of the binding domain and initiates conformational changes that serve to both activate and desensitize the channel. We assumed a set of states that are compatible with this data and constrained by additional observations: agonist binding and unbinding can occur from both closed and open states, but binding and unbinding from the open states is slow (Armstrong and Gouaux 2000), desensitization occurs from both open and closed states (Sun et al. 2002), and the unitary conductance of channels increases with receptor occupancy (Rosenmund et al. 1998; Smith and Howe 2000). These assumptions along with the observation of a double-exponential time course of desensitization (Andrasfalvy and Magee 2001) resulted in the kinetic scheme of Fig. 1A.

Setting rate constants

The independence of the subunits implies that several rate constants could be taken as multiples of the constants for single subunits, thereby reducing the number of parameters. However, since the channel needs at least two ligands to open (Swanson et al. 1997), the opening rates depend cooperatively on the number of bound ligands (Niu and Magleby 2002; Ruiz and Karpen 1999). We extended this cooperativity to the opening of the channel with three and four bound ligands (Andrasfalvy and Magee 2001) resulted in the kinetic scheme of Fig. 1A.

Setting rate constants

The independence of the subunits implies that several rate constants could be taken as multiples of the constants for single subunits, thereby reducing the number of parameters. However, since the channel needs at least two ligands to open (Swanson et al. 1997), the opening rates depend cooperatively on the number of bound ligands (Niu and Magleby 2002; Ruiz and Karpen 1999). We extended this cooperativity to the opening of the channel with three and four bound ligands, setting the ratios for the opening rates as 1:3:9. To determine the rate constants, the model response to applied glutamate was simulated using a stochastic model using standard Monte Carlo methods (Bartol et al. 1991; Franks et al. 2002; Wahl et al. 1996). Each transmitter molecule from a vesicle was tracked as it underwent diffusion through a fusion pore into the cleft. Diffusion was modeled as a random walk in three dimensions with elastic collisions at boundaries. Transmitter molecules were allowed to interact with receptor molecules and bind with a probability related to the macroscopic binding rate constant (Bartol et al. 1991). At each time step, the receptor molecules could also change their state in a probabilistic manner according to the state diagram shown in Fig. 1A. Glutamate molecules reaching the edge of the synapse at 350 nm were absorbed as a simplified model for diffusion and uptake in the neuropil surrounding the synapse (Franks et al. 2002). Because our analysis focused on the rising edge and peak of the response, we did not model the complex set of factors that determine the shape of the tail of the response.

Simulation methods

We developed a stochastic model using standard Monte Carlo methods (Bartol et al. 1991; Franks et al. 2002; Wahl et al. 1996). Each transmitter molecule from a vesicle was tracked as it underwent diffusion through a fusion pore into the cleft. Diffusion was modeled as a random walk in three dimensions with elastic collisions at boundaries. Transmitter molecules were allowed to interact with receptor molecules and bind with a probability related to the macroscopic binding rate constant (Bartol et al. 1991). At each time step, the receptor molecules could also change their state in a probabilistic manner according to the state diagram shown in Fig. 1A. Glutamate molecules reaching the edge of the synapse at 350 nm were absorbed as a simplified model for diffusion and uptake in the neuropil surrounding the synapse (Franks et al. 2002). Because our analysis focused on the rising edge and peak of the response, we did not model the complex set of factors that determine the shape of the tail of the response.
response (see Barbour 2001; Diamond and Jahr 1995; Franks et al. 2002, 2003; Rusakov and Kullmann 1998 for an analysis of these factors). As a control, we lowered the probability of absorption of glutamate at the edge such that the time constants of the decay of glutamate concentration in the cleft matched measured values (Diamond and Jahr 1997). The rise time, amplitude, and variability obtained by this modification did not differ from our standard model. We also did not include the binding of glutamate to glutamate transporters because even if such transporters are in the cleft, which is controversial, their concentration (\( \sim 100 \mu M \)) in the cleft assuming a typical density of transporters of \( 1,000/\mu m^2 \) (Lehre and Danbolt 1998) would not be sufficient to substantially affect the very high levels of glutamate present in the cleft (Fig. 2, A and B) during the early phase of the response that we have focused on. Because of the rapidity of the rising edge of the response, we suspected that it might be important to model diffusion through the fusion pore rather than approximating this as point release (Franks et al. 2003). Vesicle release was modeled as discharge through a fusion pore that formed instantaneously and expanded at a specified rate (Table 1). We found that the expansion rate of the pore over the range of 20–100 nm/ms did not affect the kinetics of the response significantly. We used a value of 50 nm/ms in all our simulations. This indicates that the point source approximation (Franks et al. 2002) is valid and could be used in cases where the computationally demanding fusion pore simulations were prohibitively slow (Fig. 3). The time step in the model (for glutamate diffusion as well as AMPA channel gating) was adaptively chosen such that value corresponding to the 95th percentile of the displacement was half the width of the fusion pore but was maximally 0.2 \( \mu s \). The Monte Carlo model was verified for different parameters (time step, unbinding steps, etc.) by comparing simulations of an instantaneous step of glutamate concentration with results obtained from deterministic equations. Additionally, we verified that the average of Monte Carlo simulations of mEPSCs matched the mEPSC calculated from deterministic equations for glutamate release through a fusion pore and diffusion through the cleft (Kleinle et al. 1996; Trommershauser et al. 1999) and the deterministic equations for AMPA channel activation.

The responses to controlled glutamate application, which we fit, were measured at room temperature (Andrasfalvy and Magee 2001). The forward glutamate binding rate for AMPA receptors has been measured to be \( \sim 10^7 M^{-1} s^{-1} \) at 5°C (Abele et al. 2000). This value was scaled up to room temperature using a temperature factor, \( Q_{10} \), of 1.4, which made it in agreement with commonly used values, e.g. (Jonas et al. 1993; Robert and Howe 2003). This value was again scaled up with the same \( Q_{10} \) to simulate EPSCs at physiological temperatures (35°C). Conformational transitions for AMPA channels are known to be more sensitive to temperature (Hestrin et al. 1990), and a higher value of \( Q_{10} \) (Table 1) was therefore used to scale up the corresponding rate constants.

An estimate of the contribution of glutamate binding to the 20–80% rise time can be obtained from dimensional considerations to be \( 1.39/k^+ [Glu] \). The factor of 1.39 is a conversion factor from the exponential time constant for the decay of free glutamate to the 20–80% rise time. This calculation assumes that binding is rate limiting in activation of the receptor compared with ligand diffusion. Figure 3, E and F, which shows multivesicular release, was generated by instantaneous release of glutamate and with a Gaussian distribution of release jitter (delay between successive release events of 150 ± 30 (SD) \( \mu s \)). It is thought that released vesicles come from a docked pool (Schikorski and Stevens 2001), which is distributed over the synapse. However, EM reconstructions show that vesicles do not overhang the edge of the synapse (Schikorski and Stevens 1997). We therefore assumed that docking sites are randomly distributed over an area starting 10 nm in from the synapse edge.

The action of cyclothiazide (CTZ) was modeled (Fig. 5) by removing all desensitized states in the model. Although there is some evidence that CTZ increases the receptor affinity to glutamate (Diamond and Jahr 1995), this was not explicitly modeled. Including this affinity increase by decreasing the unbinding rate of glutamate did not qualitatively affect the results.

### RESULTS

#### Kinetic scheme for AMPA receptors

We first constructed a model of AMPA receptor activation (Fig. 1A) to account for the activation and desensitization of channels produced by the controlled application of glutamate to excised patches at room temperature (Andrasfalvy and Magee 2001). The model builds on earlier proposals (Haussler and Roth 1997; Jonas et al. 1993) but incorporates recent evidence that channel opening depends cooperatively on the occupancy of the four subunits of the channel (Rosenmund et al. 1998; Smith and Howe 2000); the higher the number of subunits that have bound glutamate, the larger the average open channel conductance (Table 1) and the faster the opening rate. The opening rate was assumed to depend cooperatively on ligand binding (see METHODS) as has been most extensively characterized for cyclic nucleotide-gated channels (Ruiz and Karpen 1999), which like AMPA receptors have four subunits and multiple conductance states as well as Ca\(^{2+}\)-activated K\(^+\) channels (Niu and Mageleby 2002). The binding of glutamate was constrained by recent measurements (Abele et al. 2000). Based on structural observations of binding site closure due to channel opening, we assumed that binding and dissociation of glutamate from open states is much slower than from closed or desensitized states (Armstrong and Gouaux 2000; Robert and

### TABLE 1. Experimental constraints

<table>
<thead>
<tr>
<th>Observable</th>
<th>Value</th>
<th>Note</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffusion constant</td>
<td>0.4 ( \mu m^2/\mu s )</td>
<td>Free solution value is 0.76 ( \mu m^2/\mu s), reduced for tortuosity (Rusakov and Kullmann 1998)</td>
<td>Longsworth (1953)</td>
</tr>
<tr>
<td>Glutamate in vesicle</td>
<td>2000</td>
<td>200 mM</td>
<td>Burger et al. (1989)</td>
</tr>
<tr>
<td>Synapse area</td>
<td>(-0.12 \mu m^2)</td>
<td>Average mushroom spine</td>
<td>Harris and Stevens (1989)</td>
</tr>
<tr>
<td>Cleft width</td>
<td>15 nm</td>
<td></td>
<td>Wahl et al. (1996)</td>
</tr>
<tr>
<td>Fusion pore diameter</td>
<td>10 nm</td>
<td></td>
<td>(Stiles et al. 1996)</td>
</tr>
<tr>
<td>( Q_{10} ) Diffusion</td>
<td>1.3</td>
<td>Average mushroom spine</td>
<td>Matsuzaki et al. (2001)</td>
</tr>
<tr>
<td>( Q_{10} ) Rate constants</td>
<td>3</td>
<td>34°C</td>
<td>(Hestrin et al. 1990)</td>
</tr>
<tr>
<td>Conductance states</td>
<td>5, 7, and 12 pS</td>
<td>Average conductance of receptors with 2, 3 and 4 ligands bound</td>
<td>(Rosenmund et al. 1998; Derkach et al. 1999)</td>
</tr>
</tbody>
</table>
Howe 2003). The number of desensitized states was chosen to account for the double exponential time course of desensitization (Andrasfalvy and Magee 2001) and the concentration dependence of entry into desensitized states.

**Generation of the quantal response**

There is generally a single synapse on spines of CA1 pyramidal cells (Harris and Stevens 1989). This synapse has a single presynaptic active zone and a single postsynaptic density. To understand transmission at such synapses, we simulated the release of glutamate contained within a vesicle as it diffused through the fusion pore and spread within the synaptic cleft. All the determinants of this process have been estimated with reasonable certainty (Table 1). The diffusion constant of glutamate was taken to be only slightly below the rate of free diffusion (Longsworth 1953) in accord with recent arguments that this must be the case (Barbour 2001). The simulated synapse was a patch (342 nm/side, ~0.12 μm² area) corresponding in area to that of the average nonperforated synapse on mushroom or stubby spines on P15 CA1 pyramidal cells; perforated synapses are larger, whereas synapses on thin spines are considerably smaller (~0.05–0.07 μm² area) (Harris and Stevens 1989; Harris et al. 1992). Our simulations focused on mushroom spines because recent work on such spines provides a measure of the average maximal AMPA current (Matsuzaki et al. 2001), which together with the average synapse area and single-channel conductance make it possible to estimate the AMPA channel density.

The temporal and spatial pattern of cleft glutamate during the release of a single vesicle is shown in Fig. 2A and B. There is an initial “spike” of glutamate concentration in the millimolar range that extends ~100 nm from the release site. This spike is largely dissipated by diffusion within 85 μs (falling by 1/e), leaving a residual glutamate “tail” in the 100 μM range (Clements et al. 1992; Diamond and Jahr 1997). The efficacy of the spike in producing receptor occupancy is made clear by a simple calculation: for a forward binding rate of 2 × 10⁷ M⁻¹s⁻¹, glutamate at 1 mM will produce ~80% occupancy of binding sites within 80 μs (see Methods). The detailed simulations presented in the following text confirm the critical role of this spike. The importance of an analogous spike at the neuromuscular junction has been previously noted (Stiles et al. 1996).

Given the models of AMPA channel activation and of glutamate release, we then attempted to predict the properties of the quantal response. Because the density of AMPA channels at mushroom spines is now known (see preceding text), a successful model should be able to predict not only the kinetics of the quantal response but also its amplitude. Figure 2C shows the average simulated response (red trace) and the average sucrose-evoked mEPSCs (black trace) that were recorded in a proximal dendrite (Magee and Cook 2000). It can be seen that the predicted rise time (73 μs) and amplitude (11 pA) of the simulated response closely match the experimentally observed response (rise time: 71 μs; amplitude: 10 pA). The rather exact match of these quantities is probably fortuitous because there are uncertainties about the parameters in Table 1 and measurements of the quantal amplitude. A conservative conclusion would be that the mEPSC could be predicted to within about a factor of two.

**FIG. 2.** A glutamate spike in the cleft produces a hotspot of channel opening; the predicted response matches rise time and amplitude of observed responses. A: average glutamate concentration in the cleft in a disk of 100-nm radius centered at release site. B: glutamate concentration in the cleft as a function of distance from the release site at 2 different times. Elevation in glutamate concentration is spatially restricted and dilutes quickly. C: dendritic recording of miniature excitatory postsynaptic current (mEPSC) showing the rising edge (20–80% rise time is 71 μs (black curve); data from experiments described in (Magee and Cook 2000); average of 15 responses, low-pass filtered at 5 kHz). The red curve shows the mEPSC calculated from the model (20–80% rise time is 73 μs). Note that predicted responses are smoother because no recording noise was included. D: spatial distribution at the synapse (342 nm on each side) of the probability of receptor opening at response peak (average of 500 runs). E: doubling the receptor number by addition of channels at the periphery (far from fusion site) increases response by only 20%. Forty-nine channels in a 240 × 240-nm core, 100 channels in a 342 × 342-nm region (core + periphery). F: spatial distribution of the probability of receptor opening with 3 ligands bound at response peak.

Figure 2 provides further insight into the role of the glutamate spike. It can be seen that the rising phase of the response occurs during the period of the glutamate spike (Fig. 2C). Furthermore, the spatial distribution of channel opening at the peak of the response occurs in a hotspot (see also Franks et al. 2003) with dimensions similar to the spike (Fig. 2D). A measure of this hotspot is that 80% of the current is carried by channels in a 240-nm-diam region around the release site (Fig. 2E). Taken together, these results emphasize the importance of the glutamate spike in generating the quantal response. The hotspot contains ~35 channels of which about half (i.e., 15–20) are open at the peak. Of these, a substantial fraction...
(30%) has only three glutamates bound (Fig. 2F), indicating that additional glutamate binding could drive the open-state conductance to a somewhat higher value. An important consequence of the hotspot is that addition of channels distant from the site of vesicle release has little effect on quantal size (Fig. 2E). Consistent with Franks et al. (2003), we find that the quantal current varies linearly with AMPA channel density (results not shown). Because the quantal response depends more on channel density than number, quantal size will not vary strongly with synapse size or channel number at mushroom and stubby spines because the size of their synapses (0.08–0.5 μm²) is substantially larger than the hotspot (0.04–0.05 μm²) (Harris et al. 1992). However, the situation is different at “thin” spines (Harris and Stevens 1989), which have small synapses ranging from <0.04 to 0.07 μm²; the smallest of which is less than the hotspot area. For a 0.04 μm² synapse, our simulations show an average response of 6–7 pA, assuming the same receptor density as at large synapses. This must be considered an upper limit because many of the smallest synapses have lower channel density or no AMPA channels at all (Nusser et al. 1998).

As mentioned previously, estimates of quantal variation based on the width of peaks in evoked response histograms indicate that the CV is low, probably <0.2 (Stricker et al. 1996). To understand the contribution of channel opening to quantal variation, we simulated many release events at a single site, assuming a constant channel density and constant vesicle glutamate content. The computed CV was 0.18, in reasonable agreement with estimates (Stricker et al. 1996).

Properties of multivesicular evoked transmission

To understand what happens when multiple vesicles are released at single synapse, we first simulated two simultaneous vesicle fusions at fixed positions separated by 200 nm. The summation of the response to the two vesicles was nearly linear (Fig. 3A), and the channels activated by each vesicle were almost completely nonoverlapping (Fig. 3B). To gain insight into what would happen if two vesicles were released very close to each other, we released two vesicles at a separation of 25 nm, i.e., touching each other. Figure 3C shows that the average response still increased nearly linearly. The larger current arises both because the open probability and single-channel conductance increased within the hotspot and because the size of the hotspot increased (Fig. 3D). We conclude that substantial quantal summation can occur and that this is true even when vesicles are released close to each other. This is consistent with the linear increase when the glutamate concentration of a single vesicle is doubled (see following text).

To provide a more realistic picture of how multivesicular responses are generated, we simulated the situation where vesicles were released at random positions within the synapse and with a jitter in the synchronous component of transmitter release (150 ± 30 μs). The value of temporal jitter between the release of multiple vesicles is consistent with experimental measurements at other synapses (Wadiche and Jahr 2001; Wall and Usowicz 1998). Figure 3 shows that the response histograms for the release of up to two (Fig. 3E) or three vesicles (Fig. 3F) have multiple peaks. The peaks were approximately evenly spaced (Fig. 3E: fit to a sum of 2 Gaussians, mean amplitudes of 10.6 and 20.8 pA, respectively; F: fit to a sum of 3 Gaussians, mean amplitudes are 10.6, 19.8, and 28.1 pA, respectively), but the evenness was not exact. For this reason, we refer to quantal summation as quasi-linear. The width of the peaks varied (SDs = 2.5, 2.75, and 3.8 pA, respectively, with CV = 0.23, 0.14, 0.14, respectively). Thus although the responses due to vesicles released at the edge are smaller (85% of the response to on-center release), their overall contribution is not sufficient to strongly smear quantal peaks. It should be noted that this is not in contradiction with a previous report (Franks et al. 2003) arguing that position of release contributes importantly to variance because in that study, positional variability was accompanied by random variation in local receptor density, a factor assumed constant in our simulations (the
possibility of all-or-none variation in local density will be discussed later). Similarly, our results show that the jitter in the timing of vesicle release is not sufficient to eliminate peaks. We conclude that experimentally observed response histograms with multiple, sharp peaks (Stricker et al. 1996) are consistent with low variation in quantal size and could be generated at individual moderate or large CA1 synapses.

Origin of amplitude variability of mEPSCs

We next sought to understand the origins of mEPSC amplitude variability. It is this variability that has led to the idea that the CV of quantal responses is high contrary to what has been concluded through the study of evoked responses. mEPSCs have a high variability even when recorded dendritically near the site of generation by local sucrose application (Magee and Cook 2000). This variability can therefore not be attributed to differential electrotonic decrement. Importantly, these responses have different waveforms; for small and moderate mEPSCs, there is a clear correlation of rise-time and amplitude (Magee and Cook 2000). Here, we consider the basis of this size variability and the origin of the rise time/amplitude correlation.

One possibility is that the size variability is due to temporal variation in postsynaptic properties. However, this has been ruled out because responses to applied glutamate are stable over time (Liu et al. 1999). It therefore appears that variation must be due to trial-to-trial variation in cleft glutamate. Because there appears to be a small variation in vesicle diameter and because vesicle volume varies with the third power of diameter, it has been proposed that mEPSCs have substantial variation due to differences in vesicle glutamate content (Bekkers et al. 1990; Franks et al. 2003). We therefore examined how varying vesicle glutamate content affects response amplitude and found a nearly linear dependence over a fourfold range (Fig. 4A1). These results are consistent with previous simulations (Franks et al. 2002; Holmes 1995) and with experiments showing that increasing the glutamate concentration of vesicles increases response amplitude (Ishikawa et al. 2002). However, we find that increasing the vesicle glutamate content fails to account for the observed correlation of rise-time and amplitude; if anything, rise time actually becomes slightly faster for large responses (Fig. 4A2). This conclusion is in accord with recent experimental data showing that increasing vesicle glutamate content does not produce a slowing of rise time (Yamashita et al. 2003). The relevance of these findings to the basis of mEPSC variability was not previously noted. Based on these theoretical and experimental findings, it appears unlikely that mEPSCs can be attributed to variation in vesicle glutamate content. Another possible source of the observed mEPSC variability is variation in fusion pore dynamics (Choi et al. 2000; Renger et al. 2001). But mEPSCs as a result of slow pore opening would have both slow rise times and small amplitudes, contrary to the observations. Similarly, differences in electrotonic distance cannot explain the correlation because more distantly generated mEPSCs would be both smaller and slower.

A simple model that can account for the rise-time/amplitude correlation has been proposed to account for similar findings at cerebellar mossy fiber-granule cell synapses (Wall and Usowski 1998) and the thalamus (Paulsen and Heggelund 1994) and

FIG. 4. Spontaneous nearly synchronized release of multiple vesicles can account for the correlation of rise time and amplitude of mEPSCs. A1: response to different amounts of glutamate in vesicle. The response is nearly linear in glutamate concentration. A2: 20–80% rise time of response decreases with increasing glutamate, contrary to data. B1: response to release of 2 vesicles (---) released asynchronously (fixed jitter of 150 μs) at random release locations. The response to a single vesicle (—) is shown for comparison. †, the peak of the response. The 20–80% rise time of the larger response is 140 μs. B: the rise time of the response to multiple vesicles (random release jitter and random location) increases for larger responses (---: linear fit, slope = 5.4 ± 1 μs/pA; - - - : exponential fit)
is based on the idea that mESPCs can sometimes be multiquantal. Indeed, the existence of multiquantal mEPSCs in CA1 has been previously proposed on other grounds (Bolshakov et al. 1997). The explanation of the rise-time amplitude correlation is as follows: if two vesicles are released, the response will be larger because of summation and the rise time will be slower because of temporal jitter in the release. Our simulations of this model account for the observed correlation between amplitude and rise time (Fig. 4B) if we assume a jitter of $150 \pm 30 \mu s$. This model also qualitatively accounts for the fast rise times of some large mEPSCs (Magee and Cook 2000); this would happen during events when the jitter happens to be small. This explanation also accounts for observation (Magee and Cook 2000) that the largest increase in rise time occurs in the transition from one ($10 \ pA$) to two vesicles released ($20 \ pA$); when more than two are released little additional jitter occurs and there is thus little further effect on rise time.

**Functional role of desensitization**

We next sought to understand how desensitization contributes to spatiotemporal characteristics of the mEPSC. Blocking desensitization produces only a modest (1.2-fold) increase in the current at $220 \mu s$, the peak of the control response. This is in the same range (1.1- to 1.4-fold) as experimental results (Hjelmstad et al. 1999; Smith et al. 2003). In contrast, both our simulations and experiments show that blocking desensitization has a much larger effect (>2-fold) on the current during the falling phase of the response. Figure 5 gives insight into why the falling phase is particularly affected. At the peak of the response, the probability of desensitization is low but at later times, the probability increases (Fig. 5B). The lack of desensitization at the peak is because most channels that open are in the hotspot and these have three or four bound glutamates, with almost no two bound (Fig. 5A). In these states, highly liganded states, the opening rate is much faster than the desensitization rate. At later times, there are desensitized channels, and these

![Desensitization prevents slow opening of channels distant from site of release. A: spatial distribution of the probability of receptor desensitization for channels with 2–4 glutamates bound at 3 times (in $\mu s$) after the release of the vesicle. Each square is 342 nm a side as in Figs. 2 and 3. B: spatial distribution of the probability of receptor opening with 2 glutamates bound at 3 times. C, left: cyclothiazide (CTZ) increases the response to a single release (decay time-constant increased to 3.1 ms). Middle: schematic showing the release of vesicles centered on a patch of the synapse that AMPA channels. Right: CTZ increases the mean response ($n = 250$ for all cases) to an off-center (150-nm distance) release event.](image-url)
are present not only in the hotspot, but also in the periphery (Fig. 5B, middle and right) where the glutamate concentration is never high. These peripheral channels normally open with low efficiency (Fig. 5A) because a majority of them bind at most two glutamates; for these, the opening rate is slow and comparable to the desensitization rate (see Fig. 1A and Appendix A). Thus many channels distant from the release site will desensitize and not contribute to the response. If desensitization is blocked, these channels will open slowly and contribute primarily to the tail of the response.

This analysis may help to explain a surprising effect of blocking desensitization with cyclothiazide (CTZ); at some synapses with very low probability of an evoked response, CTZ can greatly increase this probability (Choi et al. 2000; Gasparini et al. 2000). To explain this finding, it has been proposed that that some vesicles release glutamate slowly (“whisper”) and that the resulting low concentration fails to activate receptors effectively unless desensitization is blocked (Choi et al. 2000). As an alternative, we suggest a postsynaptic reason for low glutamate concentration (Fig. 5C, right): synapse may be “partially silent,” i.e., contain some subregions that contain functional AMPA channels but others that do not. Vesicles released at subregions lacking functional AMPA channels will normally activate distant AMPA with such low glutamate concentration that the resulting response will be weak and below the recording noise. However, after addition of CTZ, distant release will be much more effective, producing a detectable response. The same mechanism could contribute to the increase in nEPSC frequency produced by CTZ, although presynaptic actions of CTZ are also implicated in this effect (Diamond and Jahr 1995; Yamada and Tang 1993).

Discussion

We have developed a new model of AMPA channel activation that accounts for the responses generated by controlled application of glutamate to excised patches. We then simulated the spatiotemporal pattern of glutamate concentration in the cleft and the resulting activation of AMPA channels. The predicted rise time, amplitude, and variance of quantal responses are in good accord with measured values (Fig. 2C), suggesting that the determinants of the response properties are reasonably well understood. This is the first case in which it has been possible to predict the properties of the quantal AMPA-mediated response in this way. The rather exact correspondence between simulated and measured responses is probably fortuitous given uncertainty about various parameters (Table 1). We thus estimate that the correspondence is correct within only a factor of two.

Our simulations show that most of the channels activated at the peak of the response are localized within a hotspot of 250-nm diameter centered around the site of vesicle release. The opening of these channels is driven by a brief (50–100 μs) “spike” (~1 mM) of glutamate that occurs near the site of vesicle fusion. Interestingly, the duration of the spike is close to the average time for glutamate to become unbound from the channel (1/k_{off} = 1/k_{on} = 85 μs; Appendix A), suggesting that the channel is designed to integrate over the spike duration. Because the hotspot is much smaller than moderate and large synapses, quantal size will not be expected to vary with synapse size or the total number of AMPA channels present. Rather the major postsynaptic determinant of quantal size is the density of AMPA channels. This has important implications for how synaptic strength is measured (see Conclusions and implications).

The lack of relationship between quantal size and synapse size helps to explain a surprising property of quantal transmission at the mossy fiber synapse onto a CA3 cell. At these giant spines, the presynaptic bouton makes multiple synapses of highly variable size (Sorra and Harris 1993) and number of AMPA channels (Nusser et al. 1998) but with a constant receptor density (similar to the average value in CA1). Despite this variability in channel number, histograms show narrow peaks consistent with stereotyped quantal size (Jonas et al. 1993). This finding is now quite understandable, given the importance of receptor density as opposed to number in controlling quantal size and the constancy of receptor density at these synapses (Nusser et al. 1998).

Evidence for low variation in quantal size comes from analysis of the amplitude histograms of responses evoked by minimal stimulation (stimulation of a single axon synapsing onto a target cell). These often show peaks that are narrow, indicative of low quantal variation (Bolshakov et al. 1997; Foster et al. 1991; Jonas et al. 1993; Kullmann and Nisell 1992; Larkman et al. 1991; Liao et al. 1992; Malinow 1991; Stricker et al. 1996). It has been unclear whether these peaks arise from summation at a single or multiple synapses because single axons can sometimes make multiple synaptic contacts onto the same target cell (Sorra and Harris 1993). However, several lines of evidence now indicate that multivesicular release can occur at individual synapses, suggesting that the peaks could arise due to the release of multiple vesicles at the same synapse. First, low-affinity antagonists have less effect on the response when the release probability is high than when it is low (Tong and Jahr 1994b). Second, the size of synaptically evoked Ca^{2+} transients in mushroom spines can be increased by facilitation (Oertner et al. 2002). Finally, N-methyl-D-aspartate (NMDA) and AMPA responses at single mushroom spines have high and correlated variability, consistent with the summation of varying number of quanta (Conti and Lisman 2003).

The simulations presented here support the possibility that amplitude histograms with multiple narrow peaks could arise from single synapses. We show that effective quantal summation can occur at single synapses and that fluctuation in release time and position are not sufficient to strongly smear the peaks (Fig. 3, E and F). Although summation is nearly linear, there are interactions of the glutamate released by multiple quanta that produce some nonlinearity. These may account for why the spacing between quantal peaks may not be exactly even (Jonas et al. 1993; Stricker et al. 1996). Furthermore, it has been noted (Stricker et al. 1996) that the CV of high peaks is not larger than that at small peaks as would be expected if quantal responses were independent and may even decrease and can even be smaller. This is not difficult to understand because interactions of multiple quanta can push receptor activation toward saturation, thereby reducing variation. Finally, it is important to note factors that will reduce linearity.
when currents at single synapses become high (>100 pA); these include changes in the driving force due to voltage drop across the spine neck (for a resistance of 200 MΩ (Nimchinsky et al. 2002), this could be 20 mV) and changes in the ion concentration (Qian and Sejnowski 1989).

One argument against multivesicular transmission has been that in some cases, histograms of evoked responses sometimes show wide variability but no evidence for peaks. However, this is not surprising given the fact that an axon can sometimes make multiple synapses onto the same cell (Sorra and Harris 1993). Because these can occur at different electrotonic distance from the soma, differential attenuation would obscure quantal peaks. Our general conclusion is that, at moderate and large, single CA1 synapses, multiple quanta are released and can summate quasi-linearly.

**Multivesicular mEPSCs**

It has been suggested that mEPSC variability at single synapses arises due to variation in the glutamate content of vesicles (Bekkers et al. 1990; Freking et al. 1995; Franks et al. 2003). This is based on the observed variability in vesicle diameter (Schikorski and Stevens 1997) as measured in fixed material. Assuming that none of the variability is due to differential shrinkage that occurs during fixation and that vesicle glutamate concentration is constant, it can be calculated that vesicle glutamate content would indeed be quite variable. However, both our simulations (Fig. 4) and experimental manipulations of vesicle glutamate content (Yamashita et al. 2003) show that whereas quantal size gets larger with increasing glutamate content, the rise time is either unchanged or becomes slightly faster. This is inconsistent with experimental data showing that larger mEPSCs have slower rise times. If variation in vesicle glutamate content cannot account for this fundamental property of mEPSCs, some other explanation must be sought. We show here that this correlation, and several of its detailed properties, can be simply accounted for if some mEPSCs are generated by two or more vesicles. Specifically, because vesicle release has a small temporal jitter, mEPSCs generated by two vesicles will, on average, have a longer rise time than those generated of one. However, two vesicles may sometimes be released synchronously, generating a large mEPSC with fast rise time, and this is what is observed (Magee and Cook 2000). At the other extreme, particularly long temporal jitter would lead to an observable inflection of the rising edge, which is also observed (J. Magee, personal communication). Finally, it has been previously proposed that mEPSCs can be multivesicular because their amplitude can be reduced by desynchronizing vesicle release with Sr2+ (Bolshakov et al. 1997).

Multivesicular miniature synaptic potentials have been demonstrated at other central excitatory and inhibitory synapses (Edwards et al. 1990; Korn et al. 1994; Llano et al. 2000; Paulsen and Heggelund 1994; Wall and Usowicz 1998). The strongest evidence for multivesicular mEPSCs is that their amplitude histograms show two or more peaks and that these coincide with peaks in the histogram of evoked responses. For this to be observable, cells must be electrotonically compact so that cable effects do not smear the amplitude distributions of miniature events. Such evidence has been obtained for mEPSCs in two electrotonically compact preparations, cerebellar mossy cells (Wall and Usowicz 1998) and thalamic relay cells (Paulsen and Heggelund 1994). In both cases, a correlation is found between rise time and amplitude, which is attributed to temporal jitter in the release of multiple vesicles. Finally, we note that spontaneous multivesicular release has also been observed in the neuromuscular junction (Liley 1957) but this is different from sub-quantal mPSCs (Kriebel et al. 1976), which is more controversial.

Recent work on multivesicular miniature inhibitory postsynaptic currents (mIPSCs) (Llano et al. 2000) indicates that large-amplitude mIPSCs are reduced by blocking Ca2+ release from intracellular stores, suggesting that they arise from Ca2+ “sparks.” These sparks in boutons have recently been directly visualized (Conti and Llano 2003). Presynaptic terminals at CA1 synapses have similar intracellular stores (Empetage et al. 2001), and it would now be important to test whether they have a role in mEPSC generation. To achieve tight synchronization of vesicle release, these sparks would have to have rapid onset. Very recent work in muscle cells suggests that allosterically coordinated opening of Ca2+ channels can occur on the 0.1-ms time scale (Zhou et al. 2003), which would provide tight synchronization of release. Thus the multivesicular explanation of mEPSC variability at CA1 synapses has some direct experimental support, has strong precedent at other central synapses, and is mechanistically plausible. If mEPSCs can be multivesicular, their large size variation is not inconsistent with low variation in quantal size. This allows the data on mEPSCs to be reconciled with that on evoked EPSCs, where the evidence for low variation in quantal size is strong (see preceding text). Definitive proof that mEPSCs are multiquantal at CA1 synapses would require the demonstration of multiple peaks in the amplitude histograms. However, since AMPA channel density varies from synapse to synapse in CA1 (Nusser et al. 1998), peaks would be obscured if mEPSCs are generated at a population of synapses; measurements on single synapses will thus be required.

**Relationship of synapse size to synapse function**

There is large variation (>10-fold) in the size of CA1 synapses and in their physiological properties, raising the question of whether the rules of quantal transmission will be the same at all synapses. The most abundant CA1 synapses, especially in adults, are on “thin” spines (area: ~0.05 μm2), which have few docked vesicles (Harris and Sultan 1995) and few or no AMPA channels (Nusser et al. 1998). Local glutamate uncaging indicates that many thin spines have no AMPA-mediated response but do have an NMDA-mediated response, suggesting that these spines have silent synapses (Matsuzaki et al. 2001). There appears to be a class of “nonsilent” synapses at which release is always unquantal as judged by the fact that the amplitude of nonfailure responses cannot be increased by enhancing release (Bolshakov and Siegelbaum 1995; Choi et al. 2000; Dobrunz and Stevens 1997). These synapses are probably small because similar experiments on large spines show multivesicular release (Oertner et al. 2002).

The largest CA1 synapses are on mushroom spines and seem well suited for multiquantal transmission. At these synapses, saturating current to applied glutamate can be 20–200 pA, depending on size (Matsuzaki et al. 2001; Smith et al. 2003). These values are larger than the quantal response (~10 pA),
making summation feasible. Consistent with this, a recording of the synaptically-evoked AMPA-mediated current from a single mushroom spine showed that the largest responses were ~40 pA (Conti and Lisman 2003). This was measured at the soma and so could be considerably larger at the synapse. When responses are generated by multivesicular release, some of the vesicles may be released close enough to interact in the opening of AMPA channels, while others may not. Those that are close will lead to summation of local cleft glutamate concentration, as has been inferred from experiments with low affinity antagonists (Tong and Jahr 1994a). However, even when close, effective summation of the response occurs (Fig. 3). Large mushroom spines are thus likely to have a large dynamic range.

Desensitization: the possibility of partially silent synapses

Glutamate binding sets in motion a transition to either the open state or the desensitized state; which tends to occur depends on the relative magnitudes of the rate constants. Importantly, the rate constant for opening depends cooperatively on the number of glutamates bound. For this reason, high cleft glutamate concentration favors opening, whereas low concentrations favor desensitization.

Low glutamate concentrations will occur if the transmitter is released slowly from a vesicle (Barbour 2001; Choi et al. 2003) or at a site that is distant from the AMPA channels. This idea may be of relevance to the finding that blocking desensitization increases the apparent probability of evoked responses (Choi et al. 2000; Gasparini et al. 2000) at some synapses. It has been proposed (Barbour 2001; Choi et al. 2000) that these synapses have low probability of response because many vesicles release glutamate slowly (whispering synapses); blocking desensitization allows the resulting low levels of glutamate to be effective in activating receptors. We suggest, alternatively, that synapses may have subregions that lack functional AMPA channels (partially silent synapses); vesicles released at silent regions would not open distant channels through "intra-synaptic spill-over" unless desensitization was blocked.

Conclusions and implications

It was originally thought that central synapses released only a single vesicle and that this saturated postsynaptic receptors, making transmission a binary phenomenon. The newer results discussed in the preceding text reveal that multivesicular transmission can occur and that quanta can effectively summate. This would allow the synaptic response to approach those produced by saturating levels of directly applied glutamate (Liu et al. 1999; McAllister and Stevens 2000). It thus appears that synapses have a much wider dynamic range than previously thought.

Traditionally, a synapse was described by the probability of vesicle release and by quantal size. The latter was taken as equal to mEPSC amplitude. According to the new view, to which our results contribute, these descriptions are not correct. Rather the following rules better account for the data: first, evoked transmission at moderate and large CA1 synapses can be multivesicular. The probability of release therefore cannot refer to the whole synapse but must refer to multiple components of the presynaptic active zone, possibly the number of docked vesicles (Schikorski and Stevens 2001). Second, the available evidence suggests that release is not governed by the binomial distribution (Stricker et al. 1996) probably because vesicles are not released independently as a result of fluctuations and nonstationarity of the Ca \(^{2+}\) increase evoked by action potentials (Malinow 1991; Bennett et al. 2000) Third, each vesicle released produces a stereotyped quantal response (CV < 0.2). This amplitude of the quantal response (~10 pA) at moderate or large synapses when measured close to the site of generation (Jonas et al. 1993; Magee and Cook 2000) is dependent on several postsynaptic factors including local receptor density and the single-channel conductance but does not depend on the total number of channels. Four, at moderate and large synapses, multiple quanta summate quasi-linearly to generate the evoked response. And fifth, the larger mEPSCs may be multivesicular and therefore mEPSC amplitude cannot be equated with the quantal response. Furthermore, manipulations that affect mEPSC amplitude might be acting presynaptically by affecting the fraction of mEPSCs that are multiquantal.

Understanding quantal transmission at synapses is key to understanding synaptic plasticity (Lisman 2003), but as summarized in the preceding text, traditional ways of measuring synaptic strength using mEPSC amplitude and frequency may have significant problems. Fortunately direct ways of measuring presynaptic release using FM dyes (Zakharenko et al. 2002) and the postsynaptic contribution to synaptic strength using two-photon glutamate uncaging are becoming available (Matsuzaki et al. 2001; Smith et al. 2003), which will greatly enhance our understanding of quantal transmission at synapses.

APPENDIX A

Rate constants for AMPA receptor model

The values for the kinetic rate constants at room temperature are (in s\(^{-1}\)):

\[
\begin{align*}
k_{01} &= 4 \cdot 1.43 \times 10^3 M^{-1}, k_{10} = 1 \cdot 4.00 \times 10^3, \\
k_{12} &= 3 \cdot 1.43 \times 10^3 M^{-1}, k_{21} = 2 \cdot 4.00 \times 10^4, \\
k_{23} &= 2 \cdot 1.43 \times 10^3 M^{-1}, k_{32} = 3 \cdot 4.00 \times 10^4, \\
k_{34} &= 1 \cdot 1.43 \times 10^3 M^{-1}, k_{43} = 4 \cdot 4.00 \times 10^4, \\
k_{53} &= 2 \cdot 1.43 \times 10^3 M^{-1}, k_{35} = 3 \cdot 0.00 \times 10^2, \\
k_{54} &= 1 \cdot 1.43 \times 10^3 M^{-1}, k_{45} = 4.00 \times 10^{-2}, \\
k_{43} &= 3 \cdot 1.43 \times 10^3 M^{-1}, k_{34} = 1 \cdot 4.00 \times 10^3, \\
k_{32} &= 2 \cdot 1.43 \times 10^3 M^{-1}, k_{23} = 2 \cdot 4.00 \times 10^4, \\
k_{21} &= 1 \cdot 1.43 \times 10^3 M^{-1}, k_{12} = 3 \cdot 4.00 \times 10^4, \\
\gamma_1 &= 400.0, \beta_1 = 130, \alpha_1 = 1200, \beta_3 = 130, \alpha_3 = 3600, \beta_3 = 130, \\
\gamma_3 &= 159.8, \gamma_0^1 = 8.5, \gamma_0^2 = 2 \times 159.8, \gamma_0^2 = 8.5, \\
\gamma_2 &= 3 \times 159.8, \gamma_0^2 = 8.5, \gamma_0^3 = 4 \times 159.8, \gamma_0^3 = 8.5, \\
\kappa_0^2 &= 444, \kappa_0^3 = 511, \kappa_0^4 = 1078, \kappa_0^4 = 1198, \kappa_0^4 = 2180, \kappa_0^4 = 1162, \\
\xi_1^2 &= 8.5, \xi_2^2 = 0.8, \xi_3^2 = 6.5, \xi_4^2 = 1.6, \xi_5^2 = 3.75, \xi_6^2 = 1.3, \\
\rho_1^2 &= 3943.2, \rho_2^2 = 493.4, \rho_3^2 = 2 \times \rho_2^2, \rho_4^2 = \rho_3^2, \rho_5^2 = 3 \times \rho_2^2, \\
\lambda_1^2 &= 156.8, \lambda_2^2 = 1274, \lambda_3^2 = 264.4, \lambda_4^2 = 43, \lambda_5^2 = 104.2, \\
\lambda_6^2 &= 26.4, \\
\sigma_0^2 &= 11.16, \sigma_1^2 = 0.16, \sigma_2^2 = 11.31, \sigma_3^2 = 1.05, \sigma_4^2 = 17.8, \\
\sigma_5^2 &= 1.01.
\end{align*}
\]

Acknowledgments

We acknowledge Marine Biological Laboratory for hospitality. We thank J. Magee for useful discussions and for providing the EPSC recordings and comments on the manuscript. We also thank D. Faber, C. Jahr, J. Howe, and
This work was supported by National Institute of Neurological Disorders and Stroke Grants R01-NS-35083 and R01-NS-27373. S. Raghavachari was additionally supported by a Young Investigator grant from the National Alliance for Research on Schizophrenia and Depression.

REFERENCES


Harris KM and Sultan P. Variation in the number, location and size of synaptic vesicles provides an anatomical basis for the nonuniform probability of release at hippocampal CA1 synapses. Neuropharmacology 34: 1387–1395, 1995.


Lisman J. Long-term potentiation: outstanding questions and attempted syn-
Liley AW. Spontaneous release of transmitter substance in multiquantal units.
Properties of synaptic transmission at single hippocam-
Liao D, Jones A, and Malinow R. Larkman A, Stratford K, and Jack J.
Lehre KP and Danbolt NC. The number of glutamate transporter subtype
Variability of neurotransmitter concentration
Liu G, Choi S, and Tsien RW. Magee JC and Cook EP.
Nonsaturation of AMPA and NMDA
McAllister AK and Stevens CF.
Matsuzaki M, Ellis-Davies GC, Nemoto T, Miyashita Y, Iino M, and Kasai
Malinow R. Paulsen O and Heggelund P.
Renger JJ, Egles C, and Liu G. Altered presynaptic vesicle
Zakharenko SS, Zablow L, and Siegelbaum SA.
Yamashita T, Ishikawa T, and Takahashi T.
Developmental increase in

Schikorski T and Stevens CF. Quantitative ultrastructural analysis of hip-
Schikorski T and Stevens CF. Morphological correlates of functionally
Smith MA, Ellis-Davies GC, and Magee JC. Mechanism of the distance-
dependent scaling of Schaffer collateral synapses in rat CA1 pyramidal
Sorra KE and Harris KM. Occurrence and three-dimensional structure of multiple synapses between individual radiatum axons and their target
Stiles JR, Van Helden D, Bartol TM Jr, Salperle EF, and Salperle MM.
Miniature endplate current rise times less than 100 microseconds from improved dual recordings can be modeled with passive acetycholine diffusion from a synaptic vesicle. Proc Natl Acad Sci USA 93: 5747–5752, 1996.

J Neurophysiol • VOL 92 • OCTOBER 2004 • www.jn.org