The Journal of Neuroscience, October 1992, 12(10): 3877–3895

Distributed Processing of Sensory Information in the Leech. III. A Dynamical Neural Network Model of the Local Bending Reflex

S. R. Lockery and T. J. Sejnowski

Computational Neurobiology Laboratory, Salk Institute for Biological Studies and Howard Hughes Medical Institute, San Diego, California 92186-5800

The subpopulation of identified interneurons in the local bending reflex receive multiple inputs from dorsal and ventral mechanoreceptors and have outputs to dorsal and ventral motor neurons. Their connections suggest a distributed processing mechanism in which withdrawal from dorsal, ventral, or lateral stimuli is controlled by a single population of approximately 40 multifunctional interneurons, but it is unclear whether additional interneurons dedicated to particular inputs are needed to account for each kind of bend. We therefore asked whether a model could be constructed that reproduced all behaviors without dedicated interneurons. Interneurons in the model were constrained to receive both dorsal and ventral inputs. Connection strengths were adjusted by gradient descent optimization until the model reproduced the amplitude and time course of motor neuron synaptic potentials in intracellular recordings of the response to many different stimuli. After optimization, the similarity between model and identified interneurons showed that additional dedicated interneurons are not necessary to produce all forms of the behavior. Successful optimization of networks with many fewer interneurons showed that the 40-interneuron network is redundant, raising the possibility that the interneurons have additional functions. Finally, optimizing networks with additional constraints produced better matches to some of the identified interneurons and showed that local bending can be produced by two populations of interneurons: one with outputs consistent with dorsal bending, the other with ventral bending. This suggests a simple model in which two principal types of interneurons produce many different behaviors and predicts the type of interneuron that remains to be identified.

In the leech local bending reflex, different spatial patterns of mechanosensory input lead to distinct patterns of excitation and inhibition of longitudinal muscle motor neurons (Lockery and Kristan, 1990a). For example, in response to a dorsal stimulus, the leech withdraws from the site of contact by contracting dorsal muscles beneath the stimulus and relaxing ventral muscles opposite the stimulus. Analogous motor responses are produced by ventral and lateral stimuli. These relationships constitute the input–output function of the reflex.

A previous investigation, which sought interneurons contributing to dorsal contractions in dorsal bending, identified a subpopulation of 17 interneurons (Lockery and Kristan, 1990b). Physiological determination of the inputs to these interneurons from P cells, the dorsal and ventral mechanoreceptors that provide the major input to the reflex, revealed that the majority of local bending interneurons receive excitatory input of similar, but not identical, strength from all four mechanoreceptors. This suggested that the interneurons in the reflex form a distributed representation of the spatial pattern of P cell stimulation in which the differences between input connections from P cells to interneurons produce differences in the pattern of interneuronal activation. Under this hypothesis, the different patterns of interneuron activation are translated into behaviorally correct patterns of motor output by the connections made to the motor neurons and the temporal dynamics of interneuron activity.

The complexity of the reflex makes it difficult to determine whether such a hypothesis is consistent with what is known of the local bending circuit for several reasons. First, the distributed processing mechanism must reproduce accurately the amplitude and time course of synaptic potentials recorded intracellularly in eight different motor neurons in response to eight different patterns of sensory input. Second, it must accommodate the large number of electrical and chemical synapses between local bending motor neurons. Third, this must be accomplished with a limited number of interneurons, each with realistic temporal dynamics.

To test the adequacy of the distributed processing hypothesis of the local bending reflex, we constructed a model of the circuit containing the known sensory and motor neurons, together with the expected number of interneurons, connected in a manner consistent with previous physiological results (Kristan, 1982; Granzow et al., 1985; Friesen, 1989a; Lockery and Kristan, 1990b). Connection strengths in the model were adjusted by a neural network optimization procedure (Pearlmutter, 1989) until the model accurately reproduced the amplitude and time course of synaptic potentials recorded from the motor neurons in response to eight different patterns of P cell stimulation. The optimization procedure incorporated additional physiological details, including the input resistance and time constants of each neuron, and the sign and amplitude of known connections. After optimization, interneurons in the model network closely resembled identified local bending interneurons in many respects, including the number of input and output connections, the distribution of synaptic potential amplitudes, the time course of

Received Feb. 13, 1992; revised Apr. 29, 1992; accepted May 6, 1992.

We thank W. B. Kristan, Jr., and P. Dayan for discussion. Support was provided by The Bank of America–Giannini Foundation for Medical Research, an NIH postdoctoral fellowship, The Howard Hughes Medical Institute, and The Mathers Foundation.

Correspondence should be addressed to Dr. S. R. Lockery, Computational Neurobiology Laboratory, Salk Institute for Biological Studies, P.O. Box 85800, San Diego, CA 92186-5800.

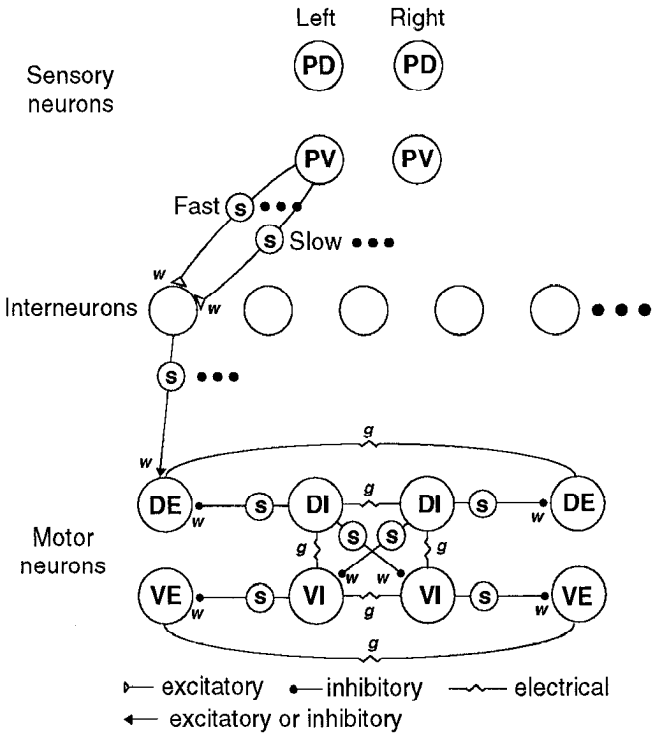


Figure 1. Network model of the local bending reflex. The network contained 4 sensory neurons (P cells), from 4 to 40 interneurons, and 8 longitudinal muscle motor neurons. Each neuron was represented as a single electrical compartment with an input resistance and membrane time constant. S-units were inserted between pairs of neurons connected by chemical synapses. The time constant of s-units represented the delays in chemical synaptic transmission. Input from sensory neurons to interneurons was mediated by parallel with fast and slow time constants. Input from interneurons to motor neurons, and between motor neurons, was mediated by single s-units with an intermediate time constant. Connection strengths (w) from sensory neurons to interneurons and from interneurons to motor neurons were adjusted by an optimization procedure to reproduce the amplitude and time course of motor neuron synaptic potentials recorded in response to stimulation of sensory neurons. Motor neurons were also connected by numerous electrical synapses (g). *PD,* P cell with dorsal field; *PV,* P cell with ventral field; *DE,* excitor of dorsal muscle; *VE,* excitor of ventral muscle; *DI,* inhibitor of dorsal muscle; *VI,* inhibitor of ventral muscle.

their synaptic potentials in response to P cell stimulation, and the effects on motor output of removing individual interneurons. The similarity between model and actual interneurons shows that a single population of interneurons generalized from the types identified so far is sufficient to account for the amplitude and time course of the motor neuron synaptic potentials. Thus, the model provides an accurate account of the local bending input–output function.

Varying the assumptions of the model showed that many other local bending networks are possible. First, changing the number of interneurons in the network revealed that accurate motor responses could be produced by a model with just four interneurons. This indicates that the biological network, which contains tens of local bending interneurons, is highly redundant. Second, constraining half the model interneurons in the complete network to represent in more detail the connectivity of the 17 previously identified interneurons, whose output is most consistent with dorsal bending, led to networks in which most of the unconstrained interneurons had outputs consistent with ventral bending. Reducing the population of unconstrained interneurons showed that no more than two are required to produce the motor neuron synaptic potentials. This places a lower bound on the number of local bending interneurons that remain to be identified and suggests a simple model in which lateral bending is produced by the combined action of two types of interneurons whose outputs are most consistent with either dorsal or ventral bending.

The dynamical model presented here generalizes a previous, static model based on a feedforward network (Lockery et al., 1989). Preliminary results based on the dynamical model are presented in Lockery et al. (1990).

Materials and Methods

Specification of the model

Circuit. The model comprised 4 sensory neurons (P cells), up to 20 pairs of left–right symmetrical interneurons, and 8 motor neurons (Fig. 1). T cells, sensory neurons responding to touch (Nicholls and Baylor, 1968), were excluded from this preliminary model, since their effects on behavior are small in comparison to the effects of the P cells (Kristan, 1982).

The maximum number of interneurons represented, in round numbers, an upper estimate of the total number of local bending interneurons in a single mid-body ganglion. A previous study (Lockery and Kristan, 1990b) identified eight paired and one unpaired interneuron contributing to dorsal local bends. In the model networks, the unpaired neuron, which is symmetrical about the midline, can be represented as an additional left–right pair, bringing the total to nine. We assumed that a search for ventral bending interneurons might yield another 9 pairs, thus 18 in total. In simulations only concerned with the general principles of local bending function, this number was rounded to 20 pairs for convenience. We refer to this as a 40-interneuron network. In other simulations, we maintained two populations of interneurons, nine pairs of dorsal bending interneurons, and either one or nine pairs of nonspecific interneurons. These we refer to as 20- or 36-interneuron networks, respectively. Pilot simulations showed that a minimum of two pairs was required to reproduce the motor neuron synaptic potentials accurately. We refer to this as a 4-interneuron network.

Each of the eight motor neurons in the model represented one of eight types of longitudinal muscle motor neurons, there being two to four neurons per type (Stuart, 1970; Ort et al., 1974). These comprise the excitors and inhibitors of dorsal longitudinal muscle, DE and DI, respectively, and the excitors and inhibitors of ventral longitudinal muscle, VE and VI, respectively. Two additional motor neurons, the L cell and cell 106, were omitted from the model because their response to the patterns of P cell stimulation used in the optimization procedure (see below) is not known.

All known chemical and electrical synaptic connections between motor neurons were included, and connection strengths (weights) were determined from previous physiological recordings (Granzow et al., 1985). Weights of feedforward connections from sensory neurons to interneurons, and from interneurons to motor neurons, were adjusted using an iterative optimization procedure (Pearlmutter, 1989) so that the model reproduced the amplitude and time course of motor neuron synaptic potentials recorded in response to single and paired P cell stimulation in eight different patterns (Lockery and Kristan, 1990a). The optimization procedure was not allowed to insert connections between interneurons, since functional connections of this type have not been found; nor was the optimization procedure allowed to insert feedback connections from motor neurons to interneurons, since only one such connection has been identified to date (Friesen, 1989b).

The previous static model excluded electrical and chemical synapses that would introduce feedback among the motor neurons (Lockery et al., 1989). Feedback requires an optimization procedure that works on networks whose activity evolves in time and can optimize neurons to follow a predetermined time course. Thus, recurrent backpropagation (Pearlmutter, 1989) was used instead of the original backpropagation algorithm (Rumelhart et al., 1986), which only applies to feedforward networks. The present model is also more realistic because the response of model motor neurons matches the real motor neurons in time course as well as amplitude. By incorporating temporal dynamics, the model is now capable of reproducing the time course of synaptic potentials in the interneurons.

Neurons. Neurons were modeled as passive, single electrical com-

partments having in parallel an input resistance (R_i) and capacitance (C) to ground. Accordingly, the rate of change of membrane potential (dV_i/dt) for neuron i was determined by the voltage V_i and the sum of the current introduced by synaptic inputs using

$$\tau_i \frac{dV_i}{dt} = -V_i + R_i I_i, \tag{1}$$

where the neuronal time constant $\tau_i = R_i C_i$, and I_i is the sum of the current introduced by chemical and electrical synapses:

$$I_i = I_i^{\text{elec}} + I_i^{\text{chem}}. \tag{2}$$

Trains of action potentials in sensory neurons were modeled as stepwise increases in presynaptic voltage. Action potentials within trains were not represented individually, since the model is concerned with the rise, fall, and amplitude of the motor neuron synaptic potentials in response to trains of sensory cell impulses, not in response to individual action potentials.

Synapses. Electrical synapses were modeled as ohmic conductances. Current introduced by electrical synapses I_i^{elec} was given by Ohm's law, and summed linearly according to

$$I_i^{\text{elec}} = \sum_j g_{ij}(V_j - V_i), \tag{3}$$

where g_{ij} is the conductance of the electrical synapse between neurons i and j. Within the ganglion, chemical synaptic transmission between inhibitory and excitatory motor neurons (Granzow et al., 1985), as well as from interneurons to motor neurons and other interneurons in the leech (Friesen, 1985; Granzow et al., 1985; Angstadt and Calabrese, 1991), is a graded function of presynaptic voltage; it does not require action potentials, nor is it substantially affected by them. The synapse between inhibitory and excitatory motor neurons was used as a model for all chemical synapses in the network, since this connection is the most thoroughly studied. In physiological experiments (Granzow et al., 1985), a stepwise increase in presynaptic voltage in DI produced a slowly rising postsynaptic potential in DE (cell 3; Fig. 2*A*). To account for the long synaptic rise time, synapse units (s-units; Fig. 1) were inserted between pairs of neurons connected by chemical synapses. The activation of each s-unit (S_{ij}) was given by

$$\tau_{ij} \frac{dS_{ij}}{dt} = -S_{ij} + f(V_j), \tag{4}$$

where τ_{ij} is the synaptic time constant and $f(V_j)$ is a sigmoidal function ($0 \leq f \leq 1$) relating pre- and postsynaptic membrane potential (see below).

The s-unit time constants combined the temporal dynamics of all the steps in transmitter release and production of postsynaptic current. Single s-units were inserted between interneurons and motor neurons, and between pairs of motor neurons. Preliminary simulations showed that the time course of motor neuron synaptic potentials, which exhibit a fast and slow exponential decay, could not be produced by a local bending model having only one s-unit with a single time constant at each connection. Therefore, two s-units, one with a fast and one with a slow time constant, were inserted between each sensory neuron and each interneuron. This was consistent with intracellular recordings of synaptic potentials from interneurons in response to P cell stimulation, which showed both fast, slow, and mixed fast and slow synaptic potentials (Lockery and Kristan, 1990b).

In the equation for s-unit activation, the sigmoidal function $f(V_j)$ represented the many processes that limit synaptic current, including limits on presynaptic calcium accumulation and number of vesicles released, the saturation of binding of transmitter to receptors, and the reversal potential of the synaptic conductance. The form of the sigmoidal function (Fig. 2*C*) was determined from published data on steady-state postsynaptic voltage as a function of presynaptic voltage at the synapses between two different inhibitory–excitatory motor neuron pairs—cells 1 and 3, and cells 2 and 4 (Granzow et al., 1985, their Fig. 5). Normalization of the extrapolated curves to maximum postsynaptic voltage showed that the form of the function was the same for these two different synapses; in each case, the data were well described by

$$f(V) = \frac{1}{1 + 77.44V^{-1.764}}. \tag{5}$$

Threshold for synaptic transmission was placed at the resting potential ($V = 0$), consistent with the fact that the inhibitory motor neurons do not release transmitter at rest (S. R. Lockery and W. B. Kristan, Jr., unpublished observations). Total current due to chemical synapses I_i^{chem} was given by

$$I_i^{\text{chem}} = \sum_j w_{ij} S_{ij}, \tag{6}$$

where w_{ij} is the strength of the chemical synapse to neuron i from neuron j, and is functionally equivalent to the product of the maximum synaptic conductance and the driving force ($E_{\text{rev}} - V$). Thus, in the model, currents from chemical synapses added linearly. Theoretical studies indicate that linear summation occurs for inputs at dendritic positions that are electronically distant (Koch et al., 1982). For closely spaced inputs, in contrast, summation can be nonlinear because the driving force changes; however, synaptic potentials in the model were kept small with respect to reversal potential. In this range, even closely spaced inputs add linearly. A more detailed model might have included reversal potentials explicitly; however, this would have greatly complicated the optimization procedure, which would have to change the reversal potential should the sign of a connection change during optimization.

Combining Equations 1–6 yields the final equation for the voltage in neuron i:

$$\tau_i \frac{dV_i}{dt} = -V_i + R_i \sum_j g_{ij}(V_j - V_i) + R_i \sum_j w_{ij} S_{ij}. \tag{7}$$

Fixed parameters

Input resistance and time constants. Motor neuron input resistance (20 MΩ) and time constant (10 msec) were estimated from the time course and amplitude of the postsynaptic potential in DE in response to a depolarizing current step (1.5 nA) injected into DI (Fig. 2*A*). Interneuron input resistance was assumed to be higher (40 MΩ) because the somata of most local bending interneurons are smaller than the somata of the motor neurons. For simplicity, the time constants for the interneurons were the same as for the motor neurons. The time constants for the s-units between motor neurons were chosen to reproduce the time course of postsynaptic voltage in DE upon injection of a square step of depolarizing current in the presynaptic motor neuron cell 1 (Fig. 2*B*). For simplicity, the same value was chosen for the s-units between interneurons and motor neurons. The same time constant was used for all s-units between interneurons and motor neurons. Time constants for the fast (10 msec) and slow (1500 msec) s-units between sensory and interneurons were fit by hand, using a simplified model network with one sensory neuron, one interneuron, and one motor neuron. The values selected provided a good fit to the time course of the intracellularly recorded motor neuron synaptic potential in response to P cell stimulation.

Electrical synapses. The coupling resistance (R_c) of electrical connections was calculated using the expression

$$R_c = \left(\frac{1}{\rho} - 1\right) R_i, \tag{8}$$

where ρ is the steady-state coupling ratio ($V_{\text{post}}/V_{\text{pre}}$) and R_i is the resistance to ground of the postsynaptic neuron excluding paths to ground provided by electrical synapses to other neurons. This expression was derived by considering the voltage divider formed by the coupling resistance in series with the shunt resistance of the postsynaptic neuron. In accordance with pairwise recordings from motor neurons (W. O. Friesen, unpublished observations; Lockery and Kristan, unpublished observations), the coupling ratio $\rho = 0.1$ was assumed for all electrical connections. This gave $R_c = 180$ MΩ for $R_i = 20$ MΩ.

Synaptic weights. Weights for chemical synapses were found by applying

$$\tau_{ij} \frac{dV_i}{dt} = -V_i + R_i w_{ij} S_{ij} \tag{9}$$

to the monosynaptic connection between inhibitory and excitatory motor neurons for the case where the maximum steady-state postsynaptic

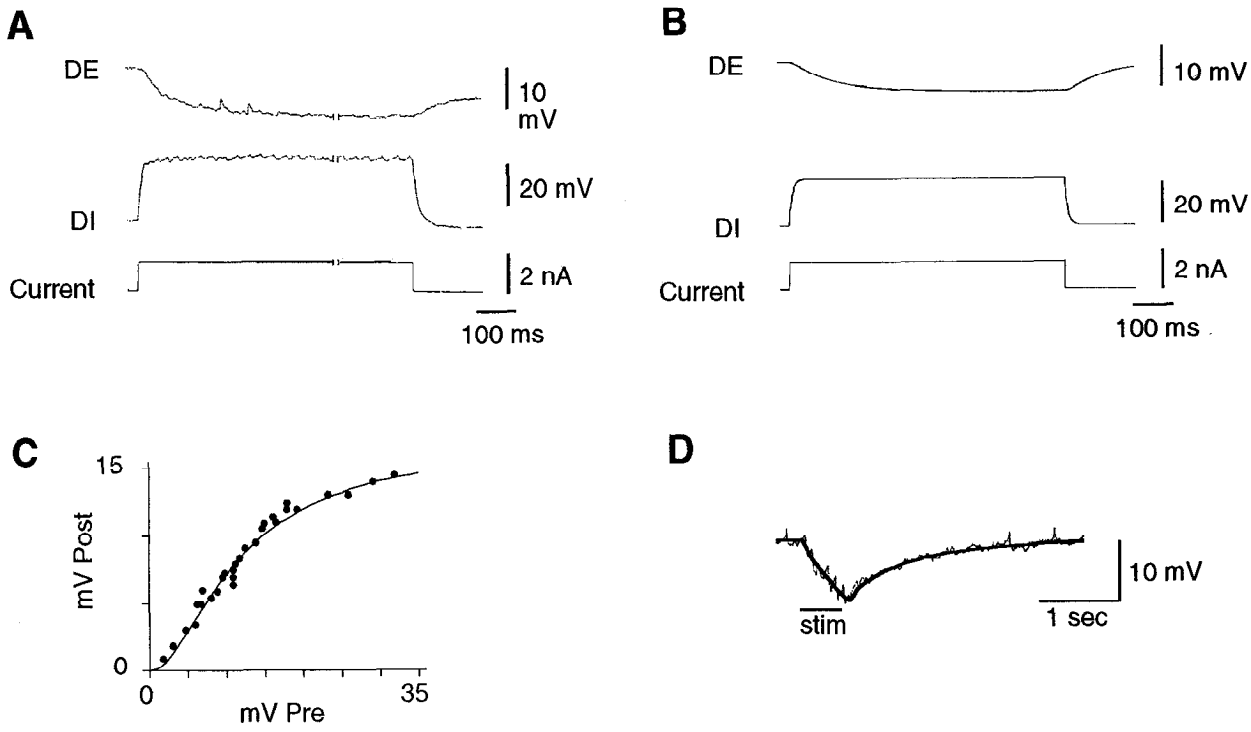


Figure 2. Physiological details of the network model. *A*, The IPSP recorded intracellularly in *DE* (cell 3) in response to depolarization by current injection of the ipsilateral *DI* (cell 1). Data are reproduced with permission from Figure 4 of Granzow et al. (1985). *B*, Simulated response of the model network to current injection in DI. Motor neuron input resistances, membrane time constants, synaptic strengths, and s-unit time constants were adjusted by trial and error to achieve the match between the recordings in *A* and *B*. *C*, Steady-state postsynaptic voltage (absolute value) in excitatory motor neurons in response to steady-state presynaptic depolarization in inhibitory motor neurons. Data points are from Figure 5 of Granzow et al. (1985). The data were fit by a sigmoidal function (see Materials and Methods) that was used for all chemical synapses in the model network. *D*, Time course of a representative motor neuron synaptic potential (*thin line*) in response to stimulation of a single P cell. The idealized curve used to construct the data set for optimization of the model network (*thick line*) overlies the physiological data. To construct the data set, the amplitude of the ideal curve was scaled to match the average peak amplitude of the synaptic potential recorded from each motor neuron in response to each of the eight patterns of P cell stimulation (Lockery and Kristan, 1990a).

voltage is achieved by maximally depolarizing the presynaptic neuron (compare Eq. 1). At maximal presynaptic depolarization, $S_{ij} = 1$ by definition, and one can solve for w_{ij}, since R_i is known. This gives the relationship

$$w_{ij} = V_i^{\max}/R_i. \tag{10}$$

$V_i^{\max}$ was estimated by fitting the sigmoidal function $f(V_j)$ to the data relating steady-state postsynaptic voltage to steady-state presynaptic voltage given in Granzow et al. (1985). For the VI to VE connection, this yielded a value of 16.7 mV; for the DI to DE connection, the value was 11.0 mV. Thus, the weight for the VI to VE connection was 0.84; for the DI to DE connection it was 0.55. The weight for the DI to contralateral VI connection was made equal to the weight for the DI to DE connection, in accordance with the data of Friesen (1989a).

Optimized parameters

Many techniques are available for optimizing weights and other parameters in neural network models. The simplest is numerical differentiation, in which a parameter is incremented by a small amount and the performance of the network is evaluated. If the performance of the network improves, the change is kept; if performance degrades, the parameter is decreased (Robinson and Arnold, 1990). Backpropagation (Rumelhart et al., 1986) relies on the same principle but evaluates all parameters simultaneously. Backpropagation is thus simply a means of finding an optimum set of parameters, much like conventional curve-fitting techniques. There is no evidence that backpropagation is used in the nervous system; it is used here only as a convenient way to search the high-dimensional space of parameters for values that are consistent with the observed behaviors.

The objective of the optimization procedure was to choose a set of connection strengths to and from the interneurons such that the time course and amplitude of motor neuron synaptic potentials in the model matched those in physiological recordings. The synaptic potential time course of a representative motor neuron recording was adopted as a template (Fig. 2*D*). The rising phase of the template was fit by

$$V(t) = 41(1 - e^{-t/1000}) + 2(1 - e^{-t/200}). \tag{11}$$

The falling phase was fit by

$$V(t) = 25e^{-t/1000} + 80e^{-t/200}. \tag{12}$$

To construct the set of input–output relations, or data set, for the optimization procedure, the amplitude of the template was scaled to the average peak amplitude for each motor neuron in response to each of eight patterns of single and paired P cell stimulation (Lockery and Kristan, 1990a, their Figs. 4, 5). The eight patterns used were (1) left dorsal P cell, (2) left ventral P cell, (3) right ventral P cell, (4) right dorsal P cell, (5) both dorsal P cells, (6) both ventral P cells, (7) left dorsal and ventral P cells, and (8) right dorsal and ventral P cells. Patterns 7 and 8 are referred to as lateral stimulation. Motor neuron response patterns associated with simultaneous stimulation of the two dorsal P cells, as well as with stimulation of the two ventral P cells, were made symmetrical about the midline by averaging corresponding responses; this was done to remove sampling error in accordance with the assumption of bilateral symmetry in the leech nervous system and behavior.

Initially, connections from sensory neurons to interneurons and interneurons to motor neurons were small and randomly assigned. In accordance with the recurrent backpropagation algorithm (Pearlmutter, 1989), connections in the network were optimized to reduce the total error (E) defined as

$$E = \sum_p \sum_k \int_0^T 0.5[V_{pk}(t) - \bar{V}_{pk}(t)]^2\, dt, \tag{13}$$

where p ranges over the input patterns in the data set, k ranges over the motor neurons in the model, and T is the duration (1 sec) of the voltage

trajectories in the simulation. The quantity $\bar{V}(t)$ is the desired voltage specified by scaling Equations 11 and 12. For each iteration of optimization, an input pattern from the data set was presented and the network's behavior was simulated for 100 time steps (10 msec/step) using forward Euler integration. The time course and amplitude of the simulated synaptic potential in each motor neuron were then compared to the desired time course for the relevant motor neuron in response to the same pattern of P cell stimulation; this gave a measure of the instantaneous error $[V_{pk}(t) - \bar{V}_{pk}(t)]$ for each motor neuron at each time step. The instantaneous error was used to calculate at each time step $(dE/dw_{ij})(t)$ the sensitivity of the error E to a change in w_{ij} at time t. This procedure was repeated for each pattern in the data set, and the total error gradient dE/dw_{ij} was calculated according to

$$\frac{dE}{dw_{ij}} = \sum_p \int_0^T \frac{dE}{dw_{ij}}(t)\, dt. \tag{14}$$

Every connection in the network was then adjusted by a small amount proportional to the total gradient

$$\Delta w_{ij} = -\epsilon \frac{dE}{dw_{ij}}, \tag{15}$$

where ϵ is the learning rate parameter. The entire process was repeated until the average instantaneous error was less than 0.18 mV. At this point in the process, an additional iteration would have reduced the error by no more than 1 part in 10,000. After 5000 or more iterations, a good match was obtained between simulated and desired motor neuron synaptic potentials (Fig. 3) for each input–output relation in the data set. The number of iterations needed was reduced by using a separate learning rate for each weight (Jacobs, 1988; Fang and Sejnowski, 1990). Optimization required 19–21 hr on an MIPS RC3240 workstation. To determine the reproducibility of optimization results, each type of network was optimized from at least six different random initial conditions. Initial input weights were chosen randomly from a uniform distribution over the interval [0, 0.1]. Initial output weights were chosen over the interval [−0.1, 0.1].

Connection strengths from sensory neurons to interneurons were constrained to be positive, since no inhibitory connections have been observed between such pairs. In addition, they were constrained to be greater than 1.35 mV to ensure that sensory input was widely distributed across all the interneurons. Because most neurons in the leech occur in left–right pairs, each model interneuron on the left was constrained to be the mirror image of a homolog on the right. This was implemented by initializing homologous weights with the same random value and updating them by the average of the amount specified by the algorithm for each weight independently (Eq. 15).

Preliminary simulations showed that the correspondence between model and actual histograms of strengths of connections from P cells depended on the amplitude of the voltage step representing the P cell stimuli that served as input to the network. Optimizing networks with different step amplitudes showed that a step of 10 mV produced the best-matching histogram (see Fig. 6*A*), and this value was adopted for all the simulations in this study. Step amplitude is a scale factor that was necessary to compensate for the absence of action potentials in the model. The amplitude of the step therefore has no biological significance.

The best correspondence for histograms of connection strengths to motor neurons (see Fig. 6*B*) was achieved when model interneurons were stimulated with current pulses large enough to bring the interneuron to motor neuron s-units into the flat region of the synaptic transfer function of Figure 2*C* (+2.5 nA, 2.6 sec). This pulse size agrees with the current pulses used in the physiological experiments to determine the actual connection strengths (Lockery and Kristan, 1990b).

Terminology

It has been standard to refer to the optimization of neural network models as training or learning. We prefer the term optimization because it connotes neither psychological nor biological events. However, in using this term we do not imply that backpropagation finds the optimum solution, because there is always the possibility of locally optimal solutions. Nor do we imply that the model is optimized for anything but the data set used in the optimization procedure.

The terms ipsilateral and contralateral, applied to sensory and motor neurons, refer to the relative positions of their receptive or projective fields, rather than to the location of their somata.

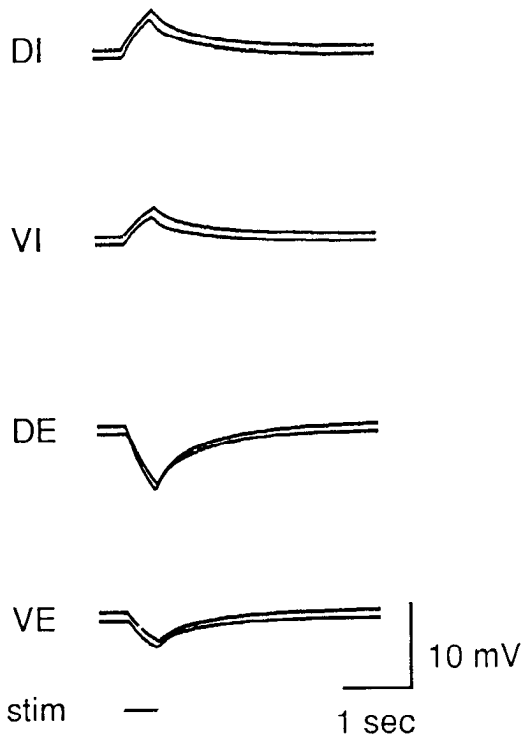


Figure 3. Simulated synaptic potentials in 4 motor neurons in the 40-interneuron model network in response to stimulation of the contralateral ventral P cell after synaptic connections had been optimized. Each panel shows the response of a single motor neuron, together with the target response from the data set. For comparison of model and target responses, the latter have been shifted upward by 1 mV. A similar match between model and target synaptic potentials was achieved for the ipsilateral motor neurons and for all other motor neurons in all other patterns.

Results

Comparison of model and actual local bending networks

Simulated injection of depolarizing current into single motor neurons in the model network reproduced the results of physiological recordings under analogous conditions (Fig. 2*A,B*). In both cases, the membrane potential in the injected neurons rose quickly to a steady-state plateau, accompanied by a slower response in the postsynaptic neuron. [The small voltage deflections in the presynaptic motor neuron (DI, Fig. 2*A*) are attenuated action potentials, which are absent in the model, because neurons were modeled as passive electrical compartments.] Thus, the input resistance as well as membrane and synaptic time constants chosen for motor neurons in the model provided a reasonable approximation to the pairwise physiological recordings.

Synaptic weights in the model from P cells to interneurons, and from interneurons to motor neurons, were chosen to optimize the correspondence between P cell–evoked motor neuron synaptic potentials in the model and physiological experiments. Comparison of motor neuron synaptic potentials in the 40-interneuron model to the desired synaptic potential wave forms showed good agreement in the amplitude and time course (Fig. 3). Figure 3 shows the response of four of the eight motor neurons to simultaneous stimulation of the contralateral P cell with a ventral receptive field; a similar degree of correspondence was found for all eight motor neurons and all eight patterns of P cell stimulation (not shown). Similar fits were obtained for the 4-, 20-, and 36-interneuron networks with the exceptions noted below.

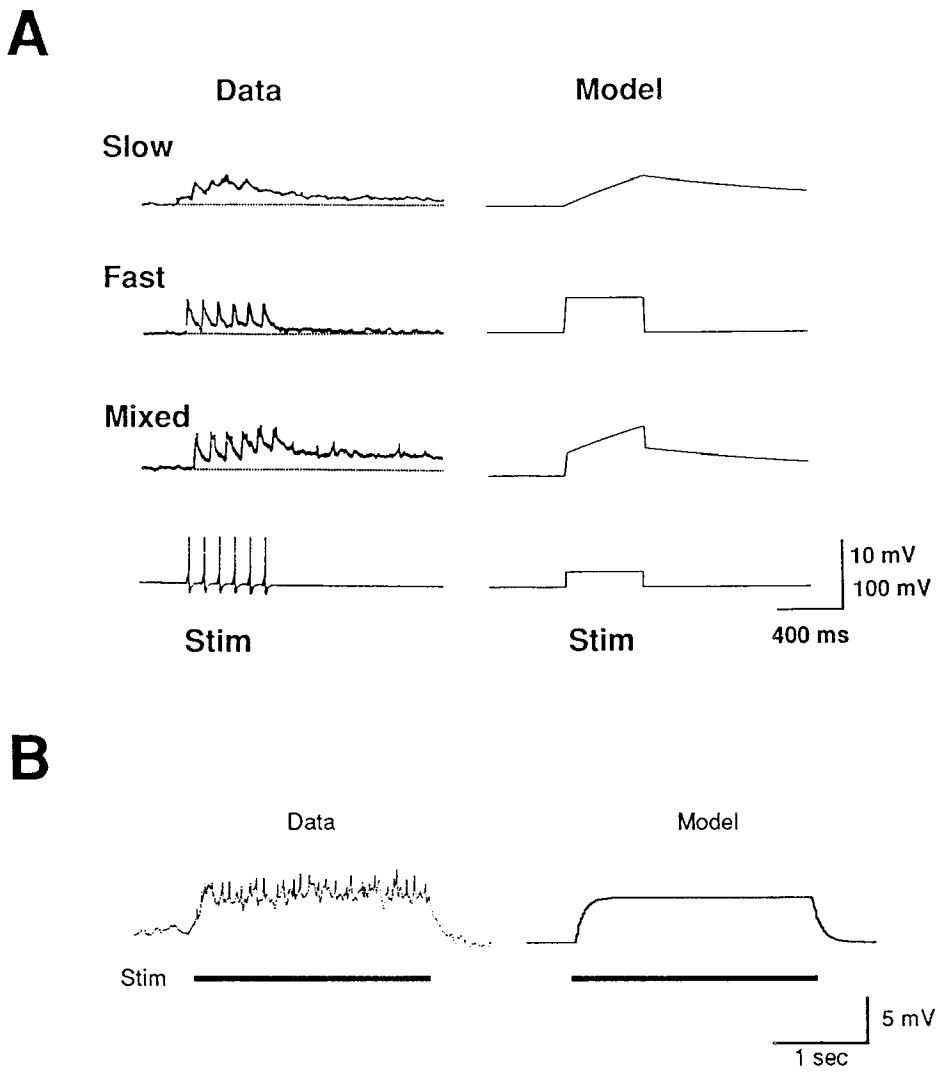


Figure 4. *A*, Comparison of recorded and simulated synaptic potentials in interneurons in response to stimulation of a single P cell (*Stim*). Physiological recordings (*Data*) are shown on the *left*, and simulations (*Model*) on the *right*. Interneurons with fast, slow, and mixed responses were seen in physiological recordings; similar types were seen in the model after optimization of the connections. In the simulation, the train of P cell impulses was modeled as a step increase in membrane potential in the sensory neuron. *B*, Comparison of recorded and simulated synaptic potentials in motor neurons in response to current injection in a single interneuron (*Stim*). The physiological recording (*left*) is from the right DE in response to a 3 nA current pulse in the contralateral cell 125. The small spikes in the motor neuron recording on the left are action potentials that originated at some distance from the somatic recording site; the model motor neuron (*right*) was represented by a passive electrical compartment that did not produce action potentials.

Because synaptic weights were chosen to optimize only the simulated motor neuron synaptic potentials, it was necessary to examine whether several other properties of the model network also matched physiological data. First, there was good agreement between simulated and recorded synaptic potentials in interneurons in response to P cell stimulation (Fig. 4*A*). Some interneurons showed mostly the slow response component, others the fast, and some showed mixed responses. Because P cell action potential trains were represented in the model as step increases in membrane potential (see Materials and Methods), summating unitary synaptic potentials were not visible in the model. Second, simulated injection of current pulses into interneurons in the model produced sustained synaptic potentials in the motor neurons, consistent with physiological recordings (Fig. 4*B*). The rise and fall of the motor neuron synaptic potential in this case were governed by the membrane time constant of the motor neuron, together with the time constant of the interneuron to motor neuron s-unit. Inspection of Figure 4*B* shows that the time to maximum response, as well as the time course of decay of the potential after stimulus offset, was adequately matched by the model.

Connectivity of each model and actual interneuron was represented as a domino with three sections, each with four squares (Fig. 5*A*). The area of each square is proportional to the synaptic strength of a single connection, with white squares for excitatory connections, and black squares for inhibitory connections. Connection strengths in the model were determined, as in physiological experiments, by finding the peak synaptic potential in the postsynaptic neuron when a standard stimulus was delivered to a single presynaptic neuron—a P cell in the case of an interneuron (10 mV, 0.5 sec), an interneuron in the case of a motor neuron (+2.5 nA, 2.6 sec). Within each domino, the top section shows connections from sensory neurons, the middle shows connections to inhibitory motor neurons, and the bottom shows connections to excitatory motor neurons. Within each section, the upper two boxes show connections from or to neurons with dorsal receptive (sensory) or projective (motor) fields; the lower two show connections to neurons with ventral fields. Left and right in the diagram correspond to left and right in the animal.

Because of the amplitude constraints on input connections (see Materials and Methods), all interneurons in the 40-interneuron model received four substantial connections from sensory neurons, though connection strength varied within and between neurons (Fig. 5*B*). Sensory input was thus widely distributed across the interneurons, capturing this salient feature of the identified interneurons (Fig. 5*A*).

No constraints were placed on the amplitude or sign of connections from interneurons to motor neurons. Therefore, to determine the correspondence between output connections of

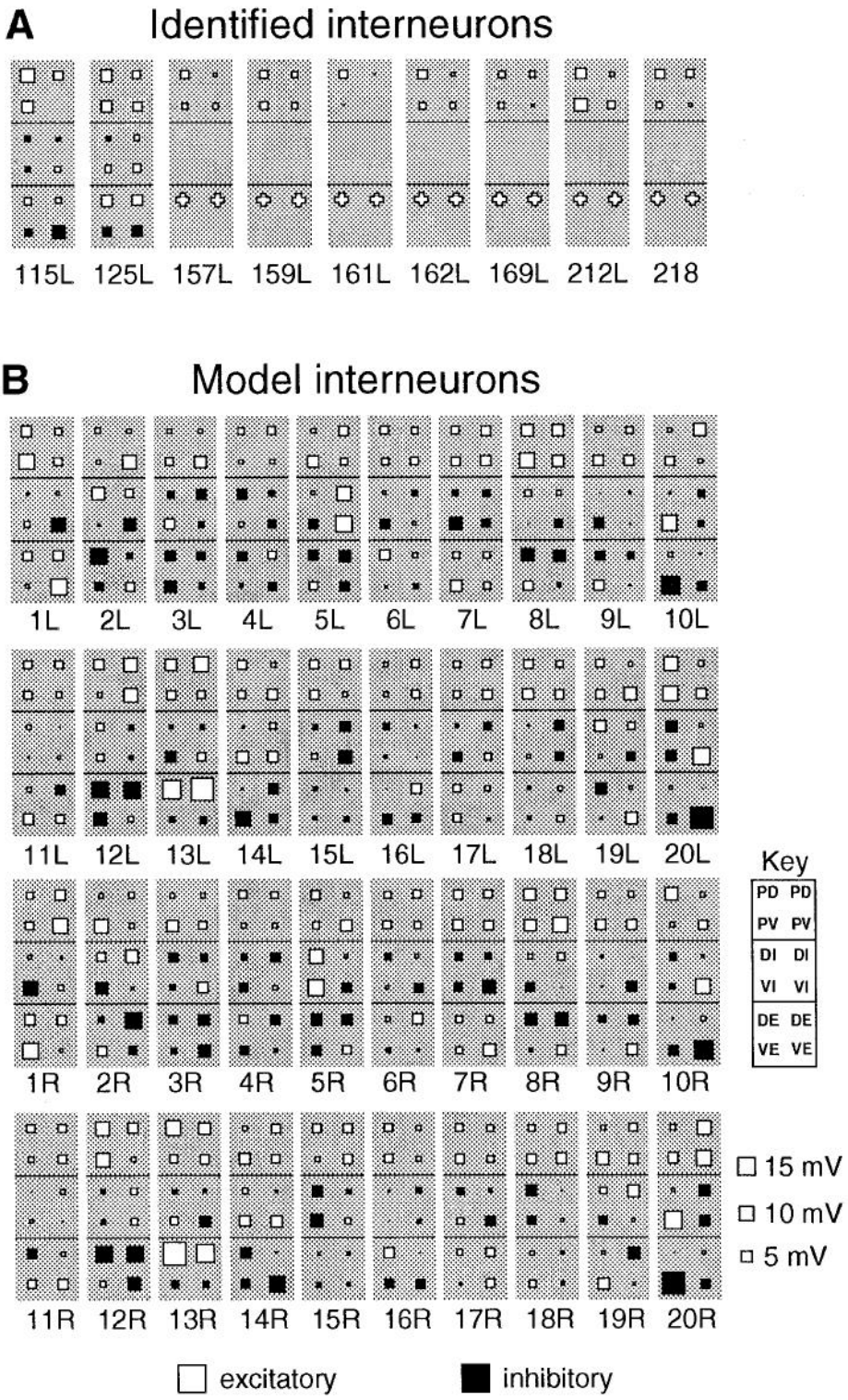


Figure 5. Connections of identified and model interneurons. *A,* Average connection strengths of identified local bending interneurons determined previously (Lockery and Kristan, 1990b). *White pluses* indicate excitatory connections of unknown strength determined from extracellular recordings of DE. *Blank spaces* indicate connections that have not been determined because the presynaptic neurons lie on the ventral surface of the ganglion while the postsynaptic neurons lie on the dorsal surface. *B,* Connection strengths of interneurons in a 40-interneuron model network after optimization. In *A* and *B,* each *square* shows the connection from the sensory or motor neuron shown in the *key. White squares* are excitatory connections; *black squares* are inhibitory connections. *Square area* is proportional to connection strength, measured as the peak synaptic potential in the postsynaptic neuron in response to a standard stimulus in sensory or interneurons. The names of identified and model interneurons are given below each interneuron (*L,* left; *R,* right). Model interneurons with the same number were constrained during optimization to be a left–right symmetrical pair in accordance with the bilateral symmetry of the identified interneurons (not shown). Like identified interneurons, model interneurons received substantial connections from four sensory neurons and most had outputs to seven or eight motor neurons.

the model and actual interneurons, counts were made of the number of interneurons having a given number of output connections above a threshold strength of 0.1 mV. The most frequent type of interneuron in the model had eight output connections, in agreement with the eight output connections of the two identified neurons for which it has been possible technically to measure output connection strengths (interneurons 115 and 125, Fig. 5*A*). Similar results were obtained for a range of thresholds, from 0.0 to 0.2 mV.

We also compared the distribution of synaptic strengths in the model and actual networks. Distributions were determined by counting the number of synaptic potentials in the actual (Fig. 5*A*) and model (Fig. 5*B*) network in 1 mV bins for input connections and 2 mV bins for output connections. For both the model and actual networks, small-amplitude input connections were more frequent than large ones and the range of input connection strengths was comparable (Fig. 6*A*). The small-amplitude output connections were also the most frequent (Fig. 6*B*),

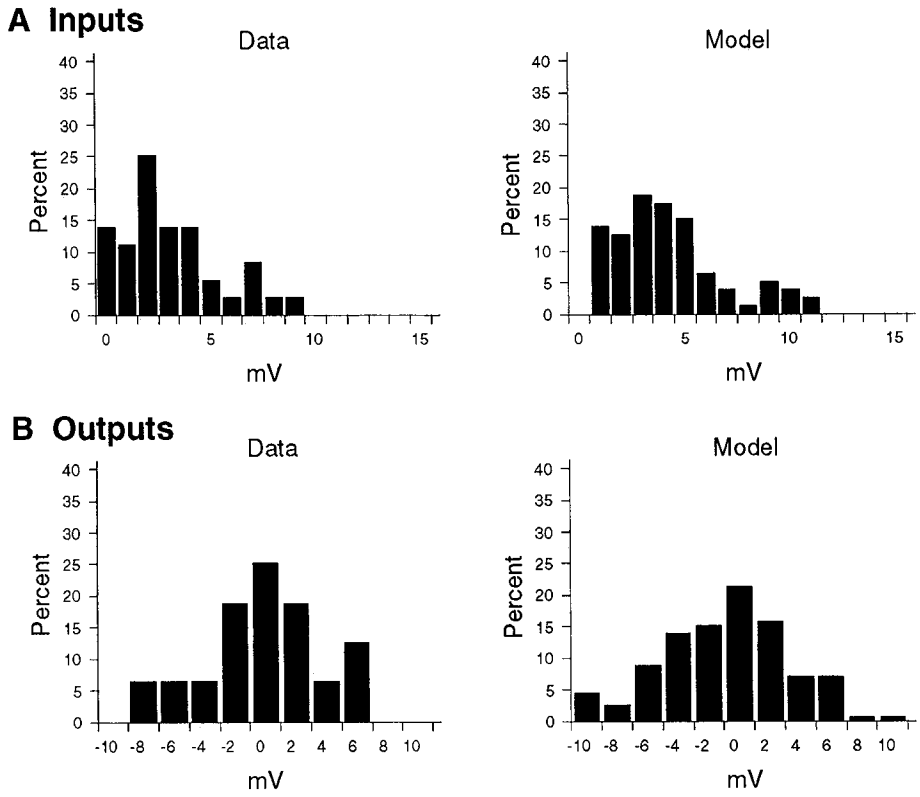


Figure 6. Histograms of synaptic potential amplitudes in interneurons (*A*) and motor neurons (*B*) in response to a standard presynaptic stimulus (see Materials and Methods). Physiological histograms are shown on the *left* (*Data*); histograms from a single model network are shown on the *right* (*Model*). The physiological data are the connection strengths shown in Figure 5*A*. The histograms showed that interneurons in the model functioned in the same operating range as in the biological network.

with both model and actual networks having approximately symmetrical distributions about zero amplitude.

The degree of similarity between model and actual interneurons, shown by the above comparisons, establishes the central result that a network composed of interneurons with left and right, dorsal and ventral inputs and multiple motor outputs is sufficient to produce the input–output function of the biological network. This means that additional types of interneurons specific for dorsal, ventral, or lateral stimuli are not required. However, further simulations showed that such interneurons can contribute to local bending in unexpected ways (see below).

Distributed representation of sensory input

Each interneuron in the model network receives input from all four P cells (Fig. 5*B*). However, inspection of the synaptic strengths in Figure 5*B* gives an incomplete picture of how sensory information is represented, and motor responses controlled, because it does not show explicitly the effect of simultaneous stimulation of pairs of P cells, or the temporal dynamics of the network. Plotting the voltage of each neuron at different times (Fig. 7) showed that each pattern of sensory input produced a different pattern of interneuron activation. For example, the pattern of interneuron activation during simultaneous stimulation of the two dorsal P cells was distinct from the pattern produced by stimulation of the two ventral P cells (Fig. 7, Dorsal vs. Ventral). Moreover, patterns changed as a function of time from stimulus onset. While almost every interneuron was strongly activated during stimulation of the two dorsal P cells (Fig. 7, Rising phase), a much smaller subset remained active after the stimulus, that is, during the falling phase of the motor neuron synaptic potentials (Fig. 7, Falling phase). A similar distinction was evident in the response to other input patterns, including stimulation of the two ventral P cells (Fig. 7, Ventral). Thus, the distributed representation of sensory information was time dependent and the falling phase of the motor neuron response was controlled by a subset of the interneurons active during its rising phase.

Tests of the necessity of interneurons for motor responses

Whether an identified interneuron made a necessary contribution to dorsal bending was tested in previous physiological experiments by hyperpolarizing the interneuron (4–5 nA) while stimulating the dorsal P cell and recording from the DE motor neuron (Lockery and Kristan, 1990b, their Fig. 7*B–D*). The resulting decrements in the DE synaptic potential were on the order of 1 mV (Fig. 8*A,C*), consistent with the large number of interneurons thought to contribute to dorsal local bending. Hyperpolarization of interneurons in the model produced effects that were similar in size (Fig. 8*B,D*).

There was also a correlation between interneuron postsynaptic potential shape and the time course of the decrement hyperpolarization of the interneuron produced in DE. Hyperpolarization of cell 125, which receives a mixed fast and slow synaptic potential from the dorsal P cell, produced a small decrement in the response of DE that was evident throughout the time course of the motor neuron synaptic potential. Hyperpolarization of mixed fast and slow interneurons in the model (5 nA) produced a small decrement like that of cell 125, which was evident throughout the motor neuron synaptic potential (Fig. 8*B*), in agreement with physiological data. Hyperpolarization of cell 218, which exhibits only the fast synaptic component, produced decrements that were visible only at the very beginning of the synaptic potential (Fig. 8*C*). Hyperpolarization of a model neuron that, like cell 218, had only the fast synaptic component produced a decrement in the rising phase of the motor neuron synaptic potential (Fig. 8*D*). In this case, there

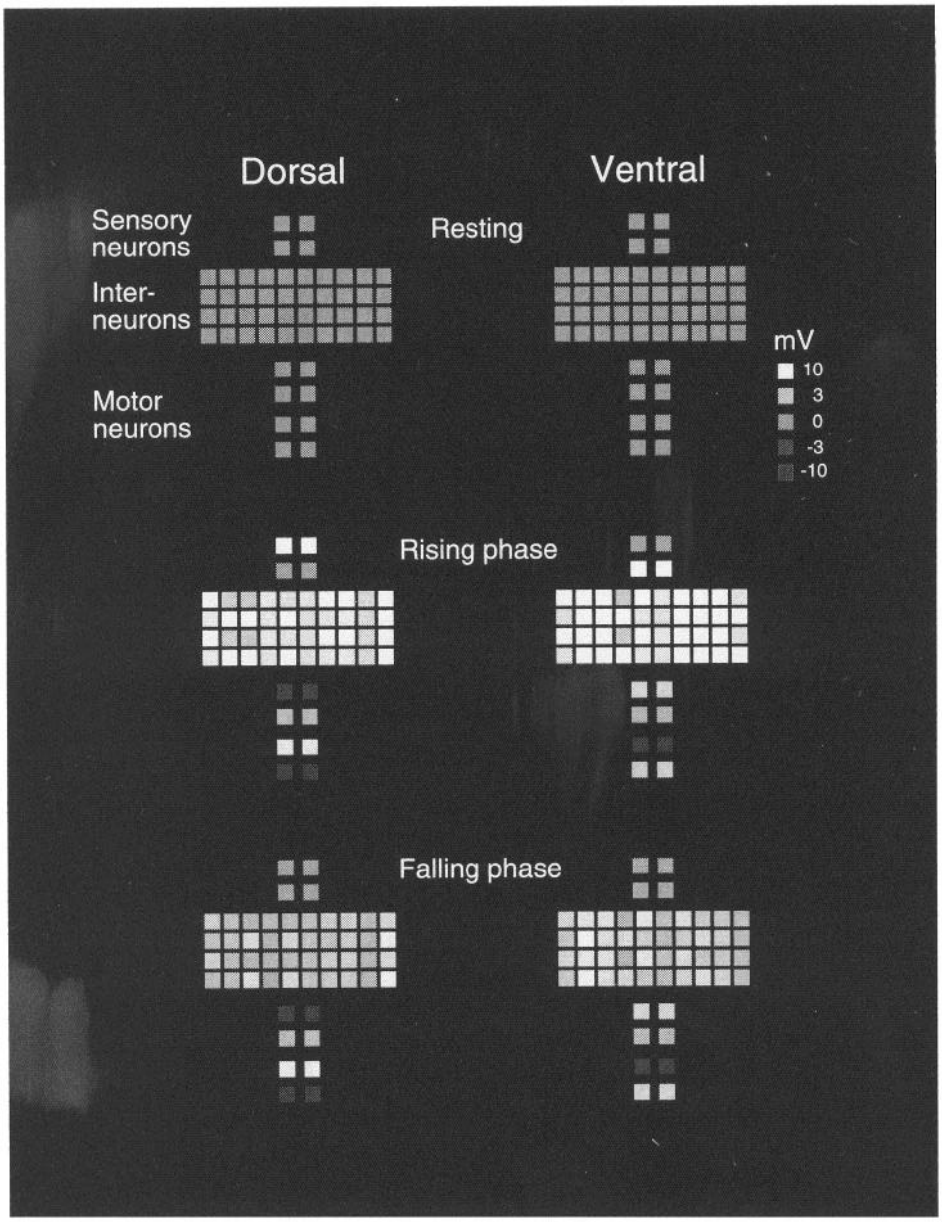


Figure 7. Response of the 40-interneuron model network to paired stimulation of the two dorsal or ventral P cells. The deviation from resting potential in each model neuron is shown at three points in time: before P cell stimulation (*Resting*), 10 msec before stimulus offset when the motor neuron synaptic potentials are still increasing (*Rising phase*), and 30 msec after stimulus offset when the motor neuron synaptic potentials are decreasing (*Falling phase*). Membrane potential is represented by the *gray scale* shown on the *right*. Different patterns of sensory input produced different patterns of interneuron activation. However, a smaller number of interneurons remain more strongly activated during the falling phase than the rising phase. Thus, the representation of sensory information varies with stimulus pattern and is time dependent.

was a discrepancy between data and model, in that the effect of hyperpolarizing the fast-type model interneuron was evident throughout the stimulus, while in the physiological experiments, it was evident only at the very beginning of the stimulus. Thus, the model can account for the amplitude but not the time course of decrements in motor neuron synaptic potentials produced by hyperpolarizing interneurons of the fast type. This suggests that in the biological system there is a process operating on a faster time scale than the fast synaptic event represented by the fast s-units (Fig. 1), and this process alters the relative contribution of individual interneurons to a given motor neuron synaptic potential. Possibilities include facilitation of the sensory to interneuron or interneuron to motor neuron synapses as well as active subthreshold conductances in the interneuron.

Sources of nonlinearity in the input–output function

An interesting property of the pattern of motor neuron activation and inactivation in response to paired stimulation of P cells is that it is not always the linear sum of the response to each P cell stimulated individually. This is evident in examining the average response patterns (Lockery, 1989), as well as in the response of individual preparations (Lockery and Kristan, 1990a). In the response of excitatory motor neurons to simultaneous stimulation of the two ventral P cells, or the lateral stimulus pattern (see Materials and Methods), this nonlinearity contributes positively to the behavior, since the excitatory motor neurons whose excitation is necessary for withdrawal from the stimulus are more excited than linear summation would predict. Because the model was optimized to reproduce accurately the average motor output to single and paired P cell stimulation, it too exhibited this nonlinearity (Fig. 9*A*). One possible source of the nonlinearity in the model is the sigmoidal function governing steady-state activation of the s-units. A second source might be the connections between motor neurons.

To determine the source of the nonlinearity in the model, a linear function was substituted for the sigmoid in a 40-interneuron network that had been optimized using the sigmoidal function (Fig. 9*D*). Almost all the motor neuron responses to

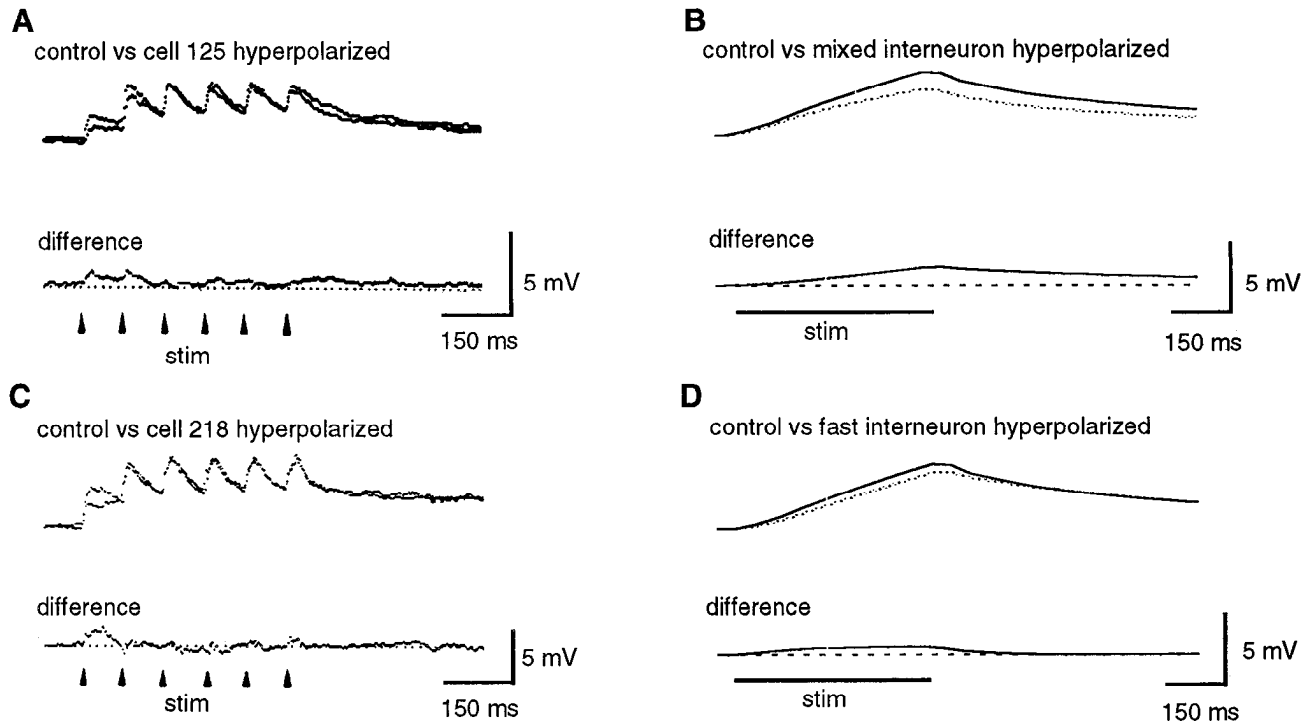
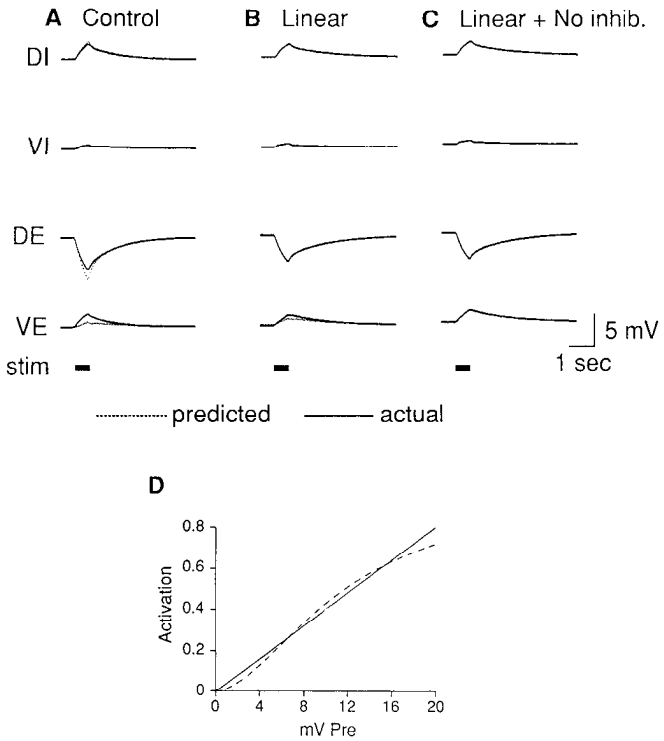


Figure 8. Contribution to motor neuron synaptic potentials of individual interneurons in the actual and model network. *A* and *C*: *Top,* Averaged intracellular recordings of EPSPs in DE in response to a train of six dorsal P cell action potentials (*arrowheads*). For one trace, the indicated interneuron was hyperpolarized (*A,* 3.8 nA; *C,* 5.1 nA). The other is a control. *Bottom,* The difference postsynaptic potential found by subtracting the hyperpolarized trace from the control. The *dotted line* is zero difference. Data are reproduced with permission from Figure 7 of Lockery and Kristan (1990b). *B* and *D*: *Top,* Simulated intracellular recordings from DE in the model in response to a standard dorsal P cell stimulus (*stim*). The *solid line* shows the control response. The *dotted line* shows the response when the indicated interneuron was removed from the model. *Bottom,* The difference postsynaptic potential found by subtracting the hyperpolarized trace from the control. The *dashed line* is zero difference. In *A* and *B,* interneurons with mixed fast and slow response components (see Fig. 4) were hyperpolarized and the decrement in DE lasts for the duration of the recording. In *C* and *D,* interneurons with only the fast response component were hyperpolarized and the decrement is transient.



paired P cell stimulation were now linear (Fig. 9*B*), indicating that the small deviations from linearity of the sigmoidal function in the model network's operating range were the major contributors to the nonlinear response to paired P cell stimulation in the model. Nonlinearities remained only in the response of VE to stimulation of the two ventral P cells, or stimulation of ipsilateral dorsal and ventral P cells (not shown). These nonlinearities disappeared when the inhibitory connection from VI to VE in the model was removed, resulting in a fully linear network

←

Figure 9. Nonlinearity in the input–output function of the local bending reflex. The *solid line* in each panel shows the simulated response of four motor neurons to simultaneous stimulation of the ventral P cells (*actual*). The *broken line* shows the response predicted by the sum of the responses to the same two P cells stimulated individually (*predicted*). *A,* Comparison of actual and predicted responses in the normal model network. Similar nonlinearities were also evident in the response to other patterns of paired P cell stimulation. *B,* Comparison of actual and predicted responses in the model network when the sigmoidal function of synaptic transmission was linearized (shown in *D*). The nonlinearities were removed, except in the *VE* motor neurons. *C,* Comparison of actual and predicted responses with the synaptic function linearized and the inhibitory connections from *VI* to *VE* removed. Actual and predicted responses now superimpose, showing that the input–output function of the network is completely linear. *D,* Linearization of the sigmoidal function governing chemical synaptic transmission. The curve shown in Figure 2*C* and reproduced here (*dashed*) was linearized by fitting a straight line to it in the range of 0–15 mV (*solid*). This range was chosen because all synaptic potentials in the model were within this range (Fig. 6*A*).

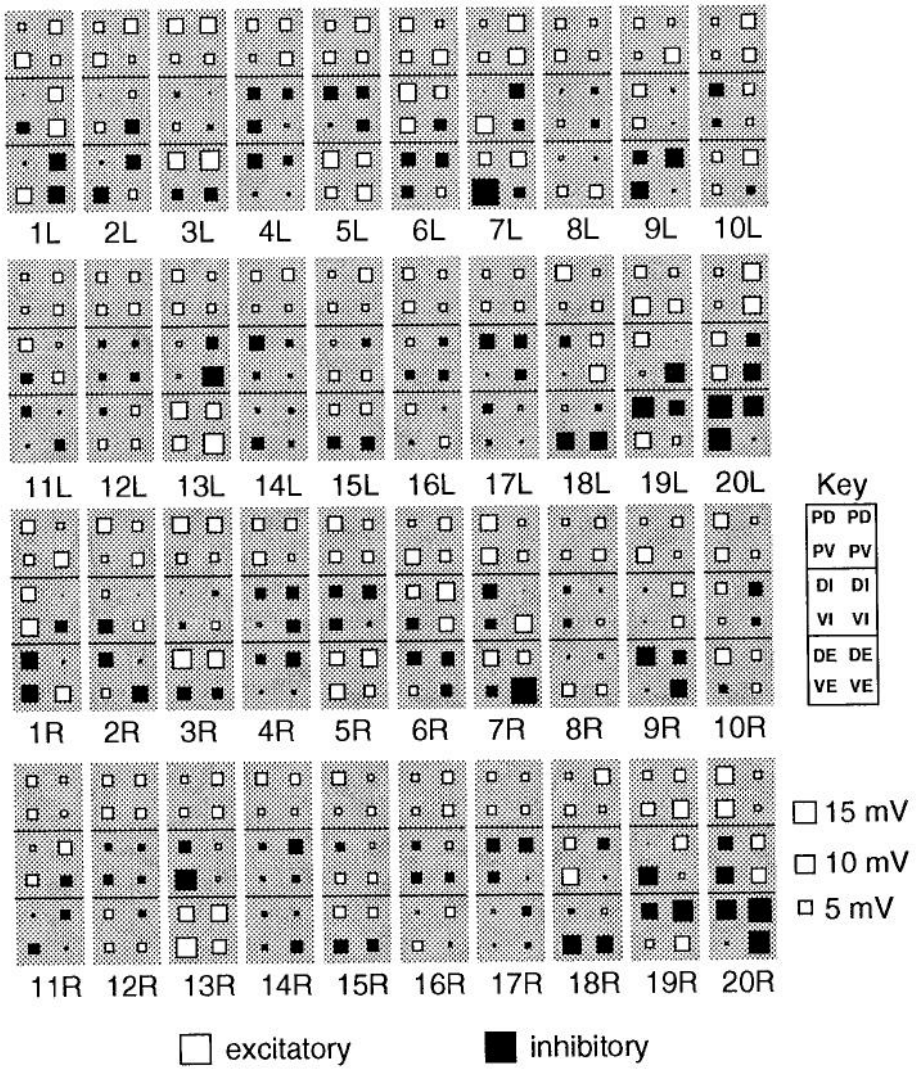


Figure 10. Connection strengths of interneurons in a second 40-interneuron model. The connectivity of this network differs from the 40-interneuron network shown in Figure 5*B* because it was optimized from a different set of random initial weights. In fact, every set of random initial weights resulted in a different 40-interneuron network. Thus, multiple local bending networks are possible. Symbols are as in Figure 5.

(Fig. 9*C*). Thus, the lateral inhibitory connections make a small additional contribution to the nonlinearity of the network.

Multiple local bending networks

To determine whether there exists a single network capable of producing the local bending input–output function or whether multiple networks exist, 40-interneuron networks were optimized from six different sets of random initial weights. Comparison of Figures 5 and 10 shows that different networks were produced by optimizing from different random initial conditions in the sense that interneuron 1L in Figure 5*B* differs from interneuron 1L in Figure 10, and so on for the other interneurons. Nevertheless, the possibility remained that by shuffling the order of interneurons, similarities between interneurons that happened to have different numbers could be found and the equivalence of superficially different networks established. For example, interneuron 7L in Figure 5*B* is similar to interneuron 5L in Figure 10. However, inspection of all interneurons in each pair of networks showed this to be the exception rather than the rule; shuffling did not reveal every neuron in one network to have an analog in the other. In dozens of 40-interneuron networks, we did not find a solution that repeated, although the statistical properties of the connections, as shown in Figures 6 and 7, were similar for all 40-interneuron networks. This result shows that there exist multiple solutions to the local bending data set.

Minimal local bending networks

To determine whether 40 interneurons are required for local bending, or whether a smaller number would suffice, we used optimization to seek solutions having fewer than 40 interneurons. As before, model networks were optimized to reproduce the time course and amplitude of motor neuron synaptic potentials in response to the eight patterns of P cell stimulation (Fig. 3). To increase the likelihood that a solution would be found, the requirement that each interneuron have an input from all four sensory neurons was removed. Networks with fewer than four interneurons could not be optimized to produce recognizable local bending motor output patterns. Therefore, the minimum number of interneurons appeared to be four. However, we cannot rigorously exclude the possibility that in the networks with fewer than four interneurons the optimization procedure became trapped in a local minimum.

In the 4-interneuron networks, final instantaneous error (0.42–0.45 mV) was higher than in networks with 40 interneurons (0.19 mV) despite much longer optimization runs (>40,000 optimization steps). Inspection of the motor neuron synaptic potentials in the minimal networks showed that although the

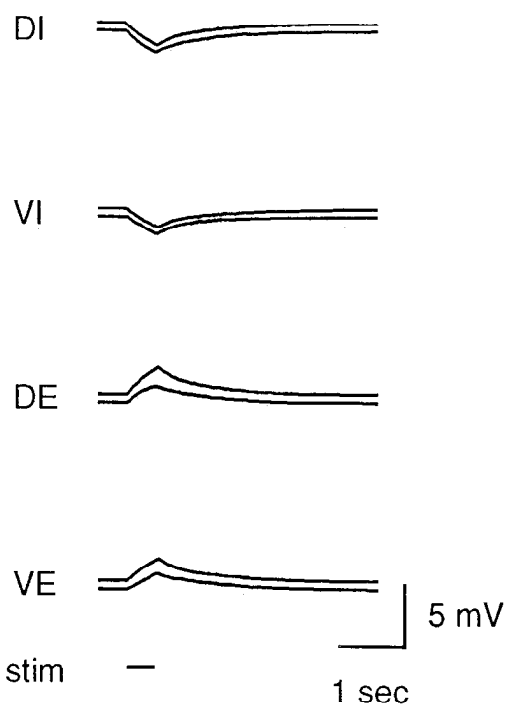


Figure 11. Simulated synaptic potentials in four motor neurons in the 4-interneuron network after optimization. Each panel shows the response of a single motor neuron, together with the target response from the data set. For comparison of model and target responses, the latter have been shifted upward by 1 mV. The model was stimulated with the lateral stimulus pattern in which the right dorsal and ventral P cells are activated. Final instantaneous error across all eight patterns in the data set was higher in the 4-interneuron network than in the 40-interneuron network, and the greatest error was in this and the other lateral pattern. Despite the higher error, the 4-interneuron model still produced local bending motor output, since in all eight patterns the polarity of the motor neuron synaptic potentials was correct.

error was increased, the polarity was correct for each motor neuron in all eight output patterns (Fig. 11). The major source of error was a reduction in the amplitude of otherwise correct synaptic potentials, particularly in lateral bending.

The 4-interneuron networks reveal several aspects of the solutions that arise when two pairs of interneurons are forced to accommodate the three basic types of local bending: dorsal, ventral, and lateral. First, interneurons differed in the degree to which they were specific for single patterns of sensory input in the data set. At one extreme (Fig. 12*A*), each interneuron responded only to dorsal or ventral stimuli. At the other extreme (Fig. 12*F*), interneurons responded to most or all inputs. Interneurons in intermediate networks (Fig. 12*B–E*) had both specific and nonspecific sensory inputs. Second, interneuron output connections revealed a strong bias toward production of a single motor response pattern: lateral local bending. In the networks shown in Figure 12, *A, B, E,* and *F,* all of the output connections (or all but one) are consistent with a pattern that produces lateral local bending. Other output types did occur, such as ventral bending interneurons (Fig. 12*C,* interneuron 2L) and approximate dorsal bending interneurons (Fig. 12*D,* interneuron 1L). However, every network had at least one pair of lateral bending interneurons.

That some networks contained nothing but lateral bending interneurons raises the question of how these networks produced dorsal and ventral bends. Inspection of the input and output connections of individual interneurons suggests that nonlateral output patterns are the result of small differences between large effects of interneurons on motor neurons. For example, in Figure 12*A,* a dorsal stimulus excites interneurons 1L and 1R. The net excitation of DEs in this motor pattern comes about because the excitatory connections to DEs are slightly stronger than the inhibitory connections. For VEs, which are inhibited in this motor pattern, the inhibitory connections are larger than the excitatory connections. A similar explanation accounts for the production of the correct net synaptic potentials in the inhibitory motor neurons if one takes the contralateral DI to VI inhibition into account (Fig. 1).

Networks with ventral bending interneurons

The previously identified interneurons (Fig. 5*A*) do not comprise the entire population of local bending interneurons. These interneurons were identified on the basis of receiving excitatory input from dorsal P cells and in turn exciting the dorsal excitatory motor neurons (Lockery and Kristan, 1990b). However, in the patterns of motor neuron synaptic potentials produced by lateral and ventral sensory stimulation, one or both of the dorsal excitors are inhibited. Because none of the identified interneurons inhibit the dorsal excitor, at least one and perhaps several types of interneurons remain to be identified.

To predict the identity of these interneurons, we divided a population of 36 interneurons in a model network into two equal subpopulations. The first subpopulation, which we refer to as dorsal bending interneurons (Fig. 13), was a generalization of the connectivity of the identified local bending interneurons in four respects. First, this group contained nine left–right homologous pairs of interneurons reflecting the eight paired and one unpaired type of identified interneuron (see Materials and Methods). Second, each interneuron in the group was constrained to receive substantial excitatory input from all four P cells, that is, both dorsal and ventral inputs. Third, connections to dorsal excitatory motor neurons were constrained to be excitatory while connections to ventral excitatory motor neurons were constrained to be inhibitory. This ensured that the output of this population would be consistent with dorsal bending, reflecting the assumption that the output connections of interneurons 115 and 125 to excitatory motor neurons are typical of the other identified interneurons. Fourth, no constraints were placed on the connections to the inhibitory motor neurons, since connections to inhibitory motor neurons from interneurons 115 and 125 are not necessarily opposite in sign to the connections to the excitatory motor neuron of the same body quadrant (Fig. 5*A*).

The second subpopulation, which we refer to as unconstrained interneurons, also contained nine left–right homologous pairs, and each interneuron was constrained to receive substantial excitatory input from all four P cells. However, no constraints were placed on the connections to either excitatory or inhibitory motor neurons. We relied instead on optimization to set these connections.

Starting from six different initial conditions, the optimization procedure found sets of connections that reproduced the local bending motor output patterns as accurately as in the networks with a single homogeneous population (not shown). The resulting dorsal bending interneurons faithfully reflected the identified interneurons, with effects on excitatory motor neurons consistent with dorsal bending and multiple effects on inhibitory and excitatory motor neurons. The most frequent type of un-

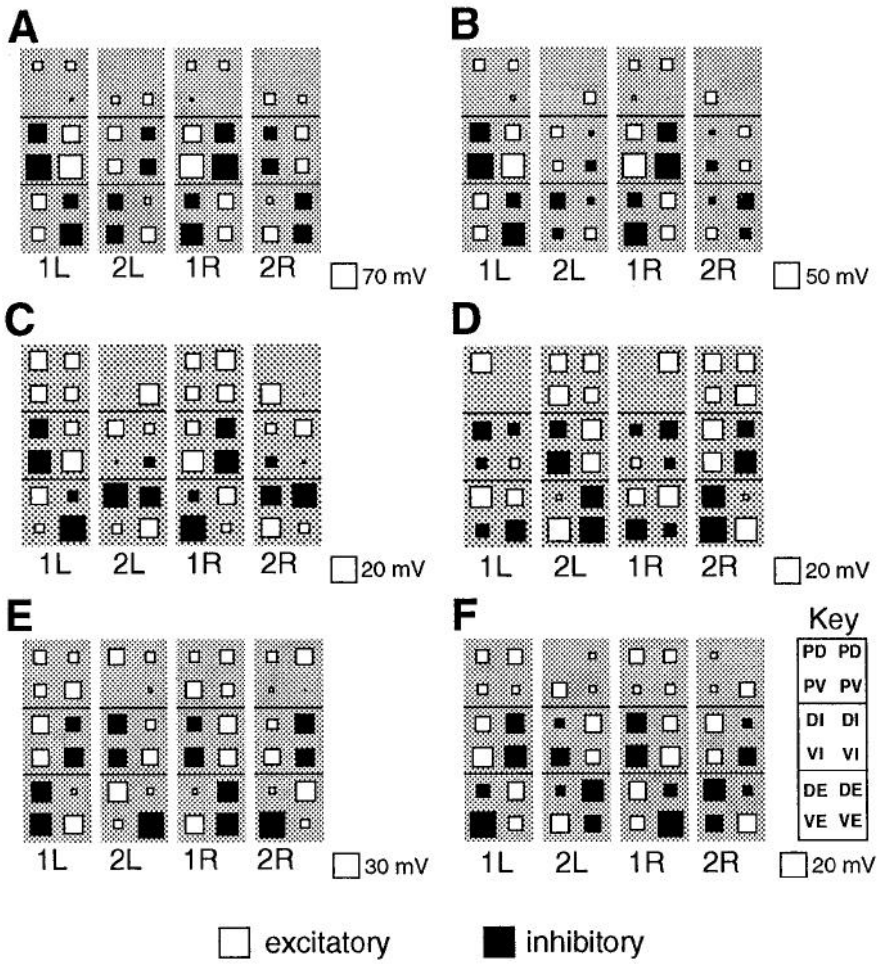


Figure 12. The range of 4-interneuron models. To determine the minimal network that was sufficient to produce local bending motor output, the number of interneurons was reduced from 40 to 4, that is, to two left–right symmetrical pairs. Smaller numbers of interneurons were insufficient, since such networks could not be optimized to produce recognizable local bending outputs. A range of solutions was produced that varied from having inputs that were specific for particular patterns of sensory input (*A*) to interneurons that responded to most or all inputs (*F*). Surprisingly, some interneurons in *A* and *B* had inputs specific for dorsal or ventral bending but outputs specific for lateral bending. In some sensory specific interneurons, there was thus a dissociation between input and output specificity. Symbols are as in Figure 5.

constrained interneuron had output connections to the excitatory motor neurons that were consistent with a role in ventral bending (Fig. 13, interneurons 10L, 11L, 12L, 14L, 17L, and their homologs). In the six networks optimized in this way, 61% of the unconstrained interneurons fell into this category. Other frequently encountered types of interneurons included ones that excited all four excitatory motor neurons (Fig. 13, interneuron 15L) and ones that inhibited all four excitatory motor neurons (Fig. 13, interneuron 16L). These categories accounted for 9% and 13% of the interneurons, respectively. The remaining 17% had unique motor effects. The predominance of ventral bending interneurons shows that a model in which the output patterns are produced by two populations of interneurons, one biased toward dorsal bending and one toward ventral bending, is consistent with the physiological details of the input–output function of the reflex. The types of interneurons that arose in the unconstrained subpopulation reflect possible connectivities of as yet unidentified interneurons. However, a different set of assumptions regarding the dorsal bending subpopulation may have led to different predictions.

That fewer than 18 unconstrained interneurons might be sufficient to produce the motor output patterns is suggested by the fact that the input–output function can be quite closely approximated by the 4-interneuron networks. This was tested by optimizing networks in which the number of unconstrained interneurons in the model was reduced from 18 to 2, that is, to a single left–right pair (Fig. 14). The resulting 20-interneuron networks reproduced patterns of motor neuron synaptic potentials with a final instantaneous error of 0.29–0.30 mV. As with the 4-interneuron network, most of the error was due to a reduction in amplitude of the synaptic potentials, especially in lateral bending. In six networks optimized from different initial conditions, the 18 dorsal bending interneurons were similar to the dorsal bending interneurons in the 36-interneuron networks, while the unconstrained pair of interneurons were like the predominant type of unconstrained interneuron in the 36-interneuron network. While there was considerable variability among the ventral bending interneurons in the 36-interneuron networks, only two types of ventral bending interneurons were seen in the 20-interneuron networks (Fig. 14*B*). In one type, one VI was excited and the other was inhibited (Fig. 14*A*); in the other type, both VIs were excited (Fig. 14*B*). Otherwise, the connections to the motor neurons were consistent with ventral bending in that the DIs were excited, the DEs inhibited, and the VEs excited. This shows that a single pair of interneurons whose effect is to produce ventral bends is sufficient to complement a population of dorsal bending interneurons qualitatively identical to the population of identified local bending interneurons. However, the output connections to the motor neurons were stronger than can be expected on physiological grounds. For example, maximum depolarization (+2.5 nA, 2.6 sec) of interneurons 10L and 10R produced IPSPs 77 mV in amplitude, which exceeds the reversal potential for IPSPs in the leech (Nicholls and Wallace, 1978; Cline, 1986). Presumably, such large synaptic effects were necessary to counteract the effect of the 18 dorsal bending interneurons. In actuality, therefore, more than

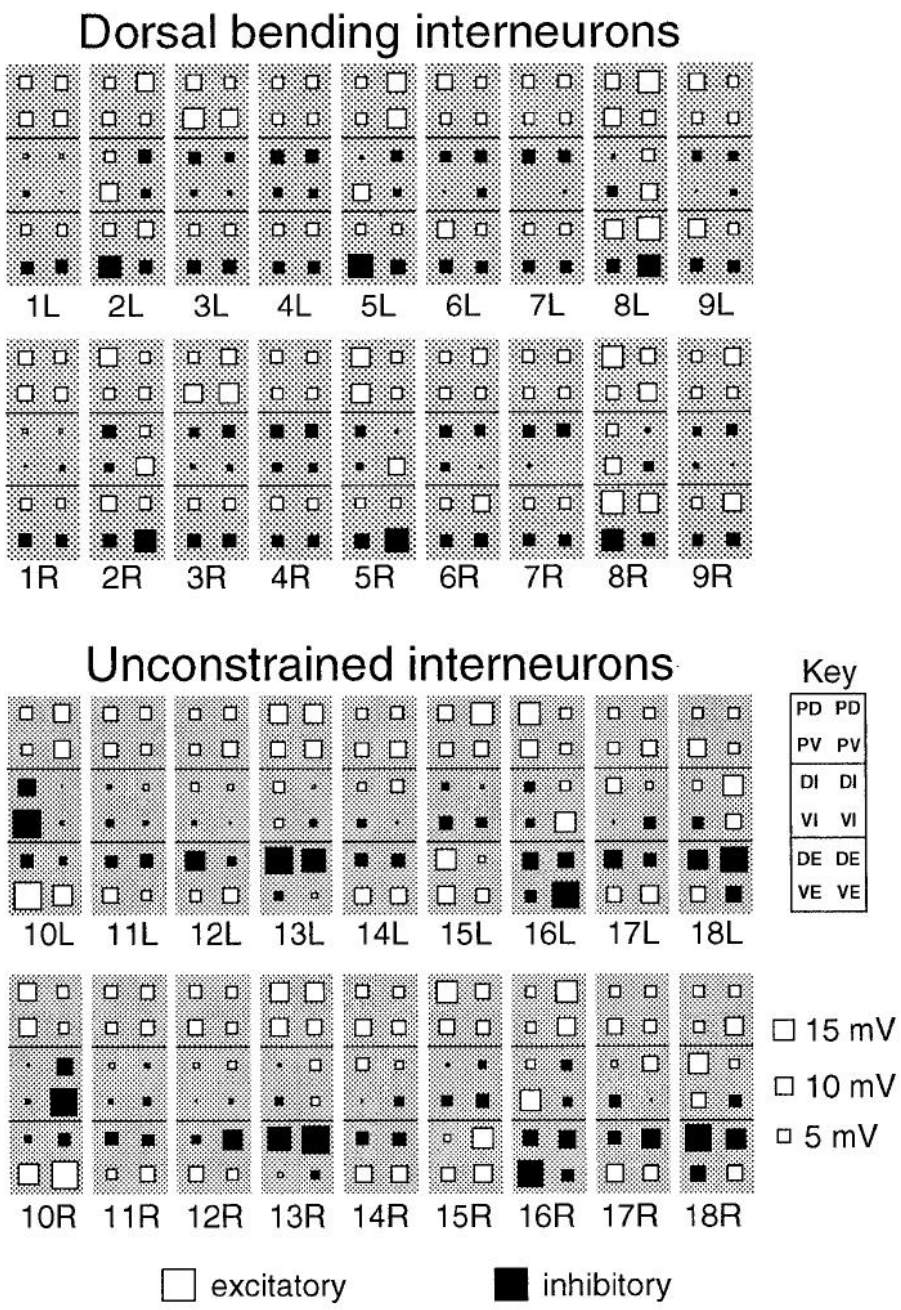


Figure 13. The 36-interneuron network model. Interneurons *1L–9R* were constrained to receive four excitatory P cell inputs and have outputs to excitatory motor neurons consistent with dorsal bending. Interneurons *10L–18R* were constrained only to receive four excitatory P cell inputs; no constraints were placed on the sign or amplitude of their output connections. After optimization to the local bending data set, most of the unconstrained interneurons had developed connections to the excitatory motor neurons that were consistent with ventral bending. Symbols are as in Figure 5.

one pair of ventral bending interneurons may be required. Nevertheless, the 20 interneuron network shows that the basic pattern of sensorimotor relations can be established with a single pair of ventral bending interneurons acting in concert with a population of dorsal bending interneurons.

Discussion

Network models of the local bending reflex

The present dynamic model confirms and extends the results of the previous static model (Lockery et al., 1989) by showing that interneurons qualitatively identical to the identified interneurons are sufficient to account for the local bending input–output function in the context of a more complete set of physiological constraints. The model now includes all the known electrical and chemical connections between motor neurons. Feedback among the motor neurons necessitated the use of recurrent backpropagation as an optimization procedure for the weights in the new model, since the original backpropagation algorithm for feedforward networks (Rumelhart et al., 1986) does not accommodate feedback connections. This had the additional benefit of including other physiological constraints, such as motor and interneuron time constants and input resistances, actual motor neuron synaptic potential time courses, and the fast and slow components of sensory neuron to interneuron connections.

Future models will incorporate additional biological constraints as more is learned about the anatomy and physiology of the reflex. The present model represents each neuron as a single electrical compartment. An important step will be to use multicompartmental neurons (Segev et al., 1989), which will

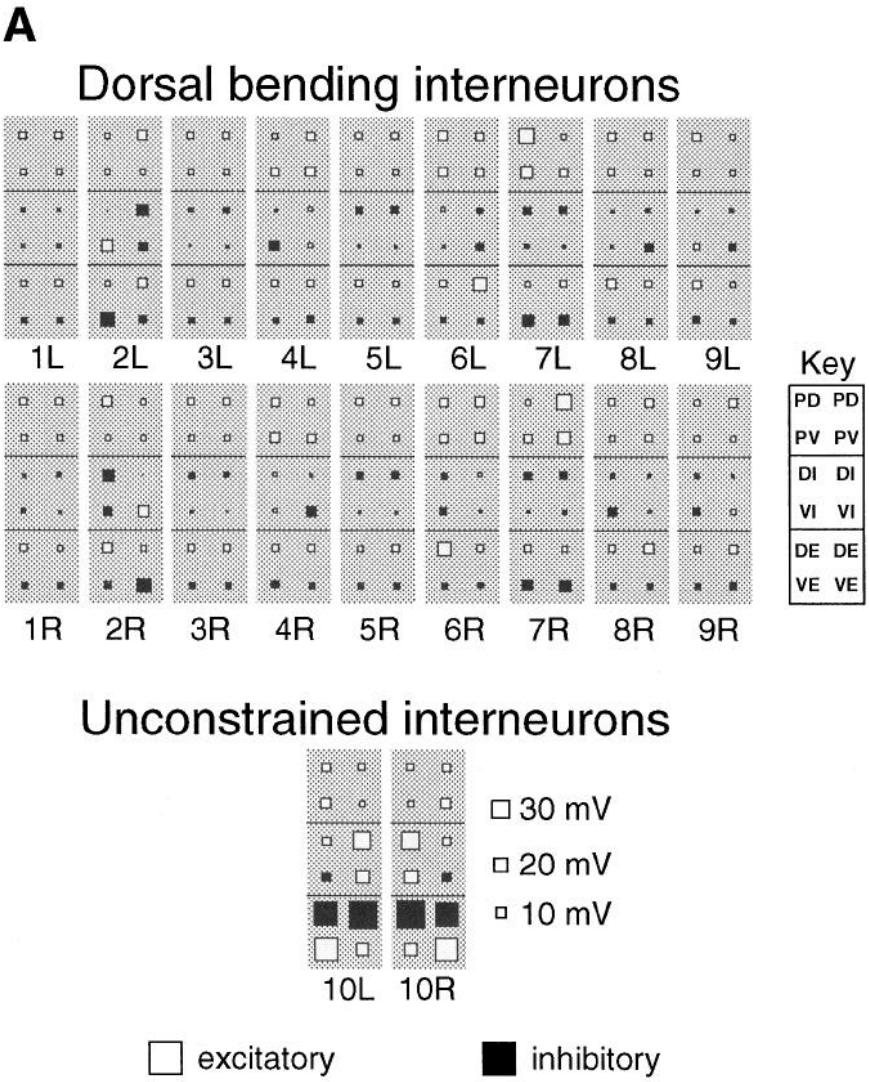


Figure 14. The 20-interneuron network model. *A*, Interneurons *1L–9R* were constrained to receive four excitatory P cell inputs and have outputs to excitatory motor neurons consistent with dorsal bending. Interneurons *10L* and *10R* were constrained only to receive four excitatory P cell inputs. After optimization to the local bending data set, this pair had output connections that were most consistent with ventral bending. *B*, Optimization of other networks from different initial conditions sometimes yielded a second type of ventral bending interneuron that excited all the inhibitory motor neurons. No other types were seen. Interneurons *1L–9R* in such networks (not shown) were similar to those in *A*. Connection strengths for the unconstrained interneurons are much larger than the dorsal bending interneurons, whose connections are shown on the same scale. The connection strengths of the dorsal bending interneurons are similar to those in the 40-interneuron networks. Symbols are as in Figure 5.

improve the accuracy of the passive properties of individual neurons. One can also add compartments representing time-dependent variables other than voltage. To some extent, this has already been done insofar as the s-units represent not voltage but the degree of activation of a synapse. Other variables include regenerative currents giving rise to action potentials or synaptic currents that undergo facilitation and depression. The latter might improve the performance of the model in tests of the necessity of interneurons for motor neuron synaptic potentials because it would enable an interneuron to make a different relative contribution to a postsynaptic response at different times during the stimulus (Fig. 8*C*).

Comparison of model and actual interneurons

After optimization, the 40-interneuron model accurately reproduced the amplitude and time course of synaptic potentials recorded in motor neurons in response to eight different patterns of P cell stimulation. Interneurons in the model resembled iden-

tified local bending interneurons in several respects. First, the time course and amplitude distribution of synaptic potentials from P cells to interneurons, and from interneurons to motor neurons, matched those of identified local bending interneurons. Second, all interneurons received four substantial input connections from P cells, and most had eight substantial output connections to motor neurons. Third, hyperpolarization of individual interneurons produced decrements in motor neuron synaptic potentials in response to P cell stimulation that were comparable in amplitude to those seen in physiological experiments. The similarity between model and actual interneurons shows that it is possible to discover a set of connection strengths between interneurons and motor neurons sufficient to resolve differences in interneuronal activation patterns associated with each pattern of P cell stimulation, and so to produce distinct patterns of motor neuron excitation and inhibition that accurately match the observed responses. It thus shows that the distributed processing hypothesis is consistent with the details of the local bending network.

The correspondence between model and actual interneurons in the 40-interneuron model was not absolute in several respects. First, no attempt was made to match the input connections of particular identified local bending interneurons. The emphasis lay instead on reproducing the general feature of having a widely distributed interneuronal representation of sensory input. Second, many of the model interneurons had connections of the same polarity to all four excitatory motor neurons, while in the two identified interneurons for which these connections have been measured, the dorsal excitors are excited and the ventral excitors are inhibited. On a more detailed level, there was never a perfect match to all eight output connections of interneurons 115 and 125 (Fig. 5*A,B*).

Discrepancies such as these are related to the functional specificities of model and actual interneurons. The identified interneurons represent a specific subpopulation contributing to the dorsal component of dorsal local bending, that is, having excitatory input from at least one dorsal P cell, and an excitatory output connection to at least one DE. Thus, the identified interneurons are functionally specific; for example, none inhibit DEs and excite VEs, connections expected of interneurons selected for contributing to ventral local bending. By contrast, the interneurons in the 40-interneuron model have no such functional specialization; rather, they combine the functionality of both dorsal bending and ventral bending interneurons, and so are different from the identified neurons in certain details of their connectivity. As more biological data become available, future models will include additional constraints on the interneurons, reducing these discrepancies.

Production of motor responses

Plotting the time course of activation of interneurons showed that they form a distributed representation of the spatial pattern of P cell stimulation in which differences between input connections from P cells to interneurons produce differences in the pattern of interneuronal activation for each of the patterns of P cell stimulation. The different patterns of interneuron activation are then translated into behaviorally correct patterns of motor neuron activation by the connections made by the interneurons onto the motor neurons and the time course of interneuron responses. In the model, the rising phase of the motor neuron response was controlled by strong activation of almost every interneuron, though which ones were most active varied across patterns. The falling phase was controlled by a subpopulation of interneurons with a sustained synaptic potential that was the result of the slow synaptic event represented by the slow s-units (Fig. 1). The similarity between model and actual interneurons in the number of sensory inputs (Fig. 5*B*) and interneuron dynamics (Fig. 4*A*) suggests that a similar functional division exists in the biological network. In fact, the divisions in the biological system may be not just between the rising and falling phase of the motor neuron synaptic potential, but also within the rising phase itself, since previous physiological data suggest that some interneurons contribute disproportionately to the very early part of the rising phase (Fig. 8*C*). Such a mechanism allows more flexibility in the construction of different time courses of motor neuron synaptic potentials.

Nonlinearity of the input–output function

We examined the source of the nonlinearity inherent in the local bending input–output function by substituting a linear approximation to the sigmoidal synaptic transfer function (Fig. 9), an experiment that is impossible to do physiologically. Almost all of the nonlinearities of the local bending input–output function in the model were attributable to the minor nonlinearities inherent in the empirical sigmoidal function used. This does not rule out the possible contribution of nonlinear summation of synaptic currents introduced by neighboring synapses on the dendritic tree of the postsynaptic neurons (Koch et al., 1982; Lytton and Kristan, 1989). Nor does it rule out a contribution from the lateral inhibitory connections between DI and DE and VI and VE. However, when the input–output function was measured, it was considered to be essentially linear (Granzow et al., 1985). It is thus significant that the rather small deviation from linearity over the working range of the model interneurons was sufficient to produce the nonlinearities inherent in the local bending input–output function. This illustrates the power of a distributed network of weakly nonlinear neurons to synthesize complex behaviors.

Multiple local bending networks

Each time the 40-interneuron model was optimized from a new set of random initial weights, a different network resulted (Figs. 5*B*, 10). Theoretically, this can be explained by the fact that there are 480 adjustable connections in the model, and only eight input–output patterns in the data set upon which the model was optimized. By analogy to a polynomial curve-fitting procedure, where the weights are the free parameters and the input–output relations are the set of points to be fit, infinitely many different high-order polynomials can be drawn through a small number of points. Despite variability at the level of individual interneurons, each of the 40-interneuron models shared such statistical properties such as the number of input and output connections and the distribution of synaptic potential amplitudes (Fig. 6). Thus, all models provided an equally good fit to the identified local bending interneurons.

Optimizing the network in the future on a larger data set will restrict the number of possible solutions, and could improve the correspondence between model and actual interneurons. The present data set contains the motor neuron responses to stimulation of single P cells and four of the six possible P cell pairs, and all P cell stimulus trains had the same firing frequency and duration. The data set could be enlarged by including responses to the remaining P cell pairs (contralateral dorsal and ventral P cells) as well as simultaneous stimulation of three and four P

cells. One might also vary the frequency and duration of the P cell stimulus trains. Finally, it is possible that some of the actual interneurons participate in several different behaviors, and have connections optimized for this multibehavioral role. In support of this possibility, interneuron 115 is known to participate in the swimming central pattern generator (Friesen, 1989b), and interneuron 125 has been shown to contribute to the shortening reflex (Wittenberg, 1991). This suggests that a better fit of the model to the data might be achieved by optimization to several different behaviors at once. Optimization can also be used to assemble models of central pattern generators (Doya and Yoshizawa, 1989; Rowat and Selverston, 1991). By optimizing a single network for local bending and swimming, it will be possible to find new ways in which single interneurons can contribute to qualitatively different behaviors.

Dedicated versus distributed local bending networks

The leech withdraws from dorsal, ventral, and lateral stimuli by contracting longitudinal muscles on the stimulated side and relaxing them on the unstimulated side. Thus, dorsal, ventral, and lateral sensory stimuli each impose a different pattern of longitudinal muscle motor neuron activation. At one conceptual extreme, this could be accomplished by four different types of interneurons, each dedicated to the detection of a particular stimulus location and the imposition of the appropriate motor pattern through a behaviorally specific pattern of output connections to the motor neurons (Fig. 15*A*). In such a dedicated interneuron network, input specificity would be provided by excitatory and inhibitory connections from the sensory neurons, and output specificity by excitatory and inhibitory connections to motor neurons.

However, optimization of 4-interneuron networks showed that even simpler dedicated interneuron networks are possible. The 4-interneuron network of Figure 12*A* has just two types of dedicated interneurons, one for dorsal and one for ventral bends, and lateral bending is produced by the concerted action of both types. It is also simpler because input specificity does not require inhibitory connections from sensory neurons to interneurons.

That the 40-interneuron model can be reduced to a model having just 4 interneurons indicates redundancy in the local bending system. Redundancy could stabilize the network's responses in the face of perturbations such as the loss of one or more connections or interneurons. Additionally, when our knowledge of the local bending reflex and other behaviors in which local bending interneurons may partake is more complete, we might find that accurate models require many more than four interneurons. These possibilities are not exclusive; the leech nervous system may have evolved for both stability and a wider range of local bending and other behaviors than the present model is currently optimized to produce.

At the other extreme, in a fully distributed network, there could be a single population of multifunctional interneurons in which each interneuron responds to all sensory inputs and the output effects of none of the interneurons are specific for a particular behavior (Fig. 15*C*). The 40-interneuron networks suggest that such a proposal is consistent with our knowledge of the input–output behavior of the reflex. However, the identified local bending interneurons, which generally receive inputs from all four sensory neurons yet have outputs that are most like the dorsal bending motor pattern (Lockery and Kristan, 1990b), are inconsistent with this view, since each interneuron responds to all sensory inputs but the output effects of individual interneurons are largely consistent with a single behavior, in this case dorsal bending (Fig. 5*A*).

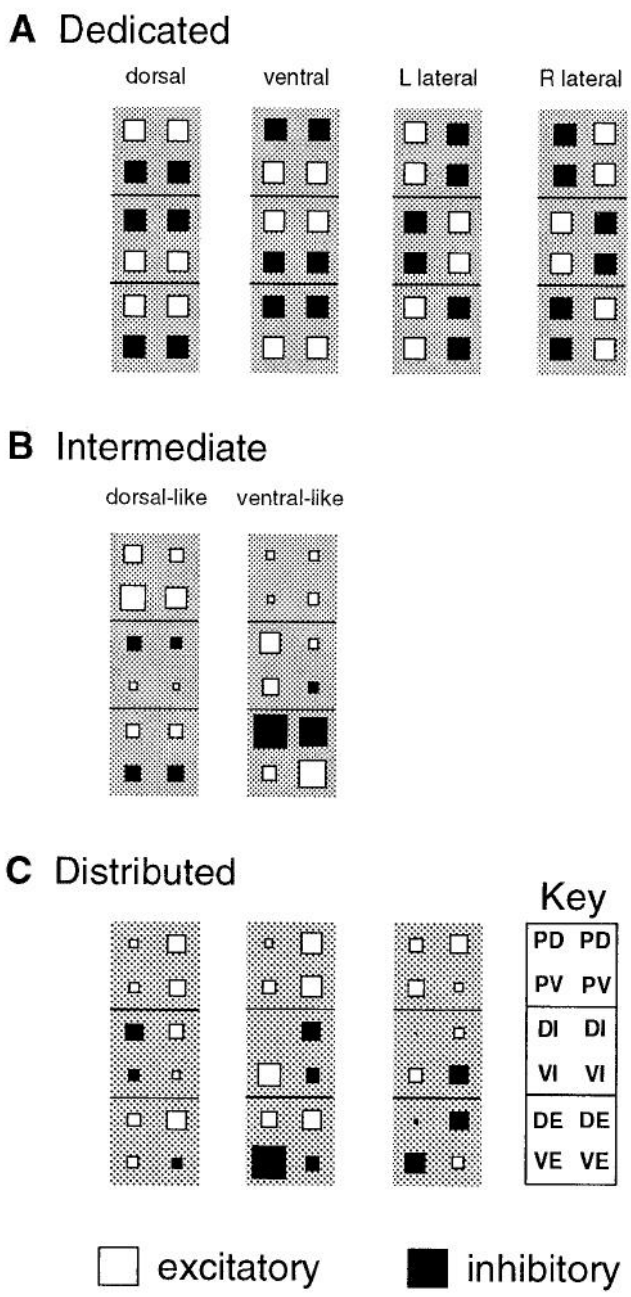


Figure 15. Three different computational strategies for the local bending reflex. *A,* Individual interneurons are dedicated to the detection of dorsal, ventral, and left (*L*) and right (*R*) lateral stimulus locations and have effects on motor output that are consistent with withdrawal from the stimulated site. These hypothetical interneurons are thus functionally specific in their inputs and their outputs. *B,* An intermediate case in which interneurons respond to all stimulus locations but have outputs that are generally consistent with withdrawal from dorsal or ventral stimuli. These interneurons are functionally specific only in their outputs. *C,* Individual interneurons respond to all stimulus locations and have outputs that are not consistent with any of the observed withdrawal responses. These interneurons are nonspecific in their inputs and outputs. The identified local bending interneurons are most consistent with the intermediate strategy, and the 20- and 36-interneuron networks show that such networks can be functional. Symbols are as in Figure 5.

The present results establish the possibility of the intermediate hypothesis in which all forms of the behavior are produced

by just two types of partly dedicated, partly distributed interneurons (Fig. 15*B*). In the 20-interneuron network, 18 interneurons were constrained to be dorsal-like, with four sensory inputs and multiple outputs to excitatory motor neurons consistent with dorsal bending: excitation of DEs and inhibition of VEs. From each random initial condition, the two unconstrained interneurons, which were also required to have four sensory inputs, developed ventral-like output connections: excitation of VEs and inhibition of DEs. This shows that all eight motor patterns, including lateral local bends, can be produced by networks with interneurons specific (or nearly so) for dorsal and ventral bends, and without interneurons specific for the lateral bending motor pattern itself. In these networks, lateral bends are produced by the summed effects of dorsal-like and ventral-like interneurons. Perhaps this represents an economical solution to local bending in which multiple types of behavior are produced by a smaller number of interneuron types.

Minimal local bending networks

The 4-interneuron networks constitute unexpected solutions to the local bending input–output function. Like the 20- and 36-interneuron networks, these had interneurons that were intermediate between the dedicated and distributed hypotheses. Although a range of solutions was produced, at one extreme these interneurons were like the ideal dedicated neurons, in that they had functionally specific inputs and outputs. Surprisingly, however, the two functional specificities were not necessarily related. For example, in Figure 12*A*, the interneurons are specific for dorsal or ventral stimuli but lateral withdrawal behavior. At the other extreme, interneurons responded to many or all inputs, yet had outputs consistent with a single behavior (e.g., Fig. 12*F*). Interneurons in the 4-interneuron networks are thus deceptive in that they have a wider functional role than their input and or output specificity would suggest, since each network produces dorsal, ventral, and lateral responses. This draws attention to the pitfalls, in distributed processing networks, of interpreting the function of single neurons in isolation.

The connectivity of unidentified local bending interneurons

The networks in which unconstrained interneurons were allowed to develop alongside a subpopulation constrained to represent the 17 identified local bending interneurons predict aspects of the connectivity of unidentified local bending interneurons. The 20-interneuron model predicts that the net effect of all the unidentified interneurons is to produce a motor output pattern that resembles ventral bending, since the two unconstrained interneurons must perform the function of the entire subpopulation of unidentified cells, however many there may be in the biological network. When the number of unconstrained neurons was increased to 18 in the 36-interneuron model, the predominant type of interneuron was still ventral-like. We propose, therefore, that a prominent feature of the unidentified local bending interneurons will be output effects biased toward ventral bending and that lateral bending is the result of simultaneous activation of dorsal and ventral bending interneurons. However, we cannot exclude the possibility that the unidentified interneurons might include lateral bending interneurons. The solutions found in the 20- and 36-interneuron networks may simply be the local minima nearest the random initial starting point and other more distant minima might contain lateral bending interneurons. We nevertheless favor the dorsal–ventral hypothesis because it accounts for all forms of local bending with two basic interneuron types and is thus more parsimonious than postulating additional lateral bending interneurons.

This hypothesis can be tested by attempting to locate and identify other local bending interneurons. Interneuron 115, which has dorsal-like outputs in local bending (Fig. 5*A*), is active during the phase of swimming in which dorsal contraction occurs (Friesen, 1989b). Thus, a precedent exists for a relationship between the phase of an interneuron in the swim central pattern generator and a role in a particular form of local bending. Perhaps swim oscillator neurons active during or near the ventral phase of the swimming motor pattern, including cells 28, 60, 33, and 27 (Friesen, 1989b), also serve as ventral bending interneurons. This could be tested by determining P cell input and motor output connections of these interneurons.

Local bending as a model system in computational neuroscience

The primary function of interneurons in the local bending network is to associate with each sensory stimulus that pattern of motor neuron excitation and inhibition that is required to withdraw from the site of contact. In computational terms, the reflex computes a function between a four-dimensional input vector encoding stimulus location and an eight-dimensional output vector encoding the associated movement. Because most interneurons in the reflex receive inputs from all four sensory neurons, the computation is achieved using a distributed representation of sensory input and motor output. In form and function, the reflex thus bears strong resemblance to perceptrons (Churchland and Sejnowski, 1992) and other artificial neural networks that excel in many biologically relevant tasks including pattern recognition, data compression, interpolation, signal detection, and prediction. The present study shows that local bending networks can vary greatly in the number of interneurons used and the degree of sensory and motor specificity. Thus, the actual local bending circuit must be one among many biologically plausible networks with identical motor output. Based on the connectivity of the identified local bending interneurons, it appears that the actual local bending network has adopted a computational strategy that, though distributed, involves aspects of a solution using dedicated interneurons. The 36-interneuron network shows that such a circuit can produce local bending responses. The question whether the biological system actually functions in this way will require identification of additional local bending interneurons and measurement of their input and output connection strengths. The question why the leech nervous system has adopted this solution will require a more complete definition of the behavioral constraints under which the interneurons operate, including the response of the motor neurons to sensory inputs that are spatially and temporally complex, and the role the interneurons may have in other behaviors that involve the same motor neurons such as swimming, shortening, and stepping (Kristan et al., 1988). That the reflex occurs in a comparatively simple animal with an easily accessible nervous system composed of identifiable neurons provides a unique opportunity to understand completely a real-life example of powerful and potentially quite general computational mechanisms.

References

Angstadt JD, Calabrese RL (1991) Calcium currents and graded synaptic transmission between heart interneurons of the leech. Neuroscience 40:185–234.

Churchland PS, Sejnowski TJ (1992) The computational brain. Cambridge, MA: MIT.
Cline HT (1986) Evidence for GABA as a neurotransmitter in the leech. J Neurosci 6:2848–2856.
Doya K, Yoshizawa S (1989) Adaptive neural oscillator using continuous-time back-propagation learning. Neural Networks 2:375–385.
Fang Y, Sejnowski TJ (1990) Faster learning for dynamic recurrent backpropagation. Neural Comput 2:274–282.
Friesen WO (1985) Neuronal control of leech swimming movements: interactions between cell 60 and previously described oscillatory neurons. J Comp Physiol A 156:231–242.
Friesen WO (1989a) Neuronal control of leech swimming movements. I. Inhibitory interactions between motor neurons. J Comp Physiol A 166:195–203.
Friesen WO (1989b) Neuronal control of leech swimming movements. II. Motor neuron feedback to oscillator cells 115 and 28. J Comp Physiol A 166:205–215.
Granzow B, Friesen WO, Kristan WB Jr (1985) Physiological and morphological analysis of synaptic transmission between leech motor neurons. J Neurosci 5:2035–2050.
Jacobs RA (1988) Increased rates of convergence through learning rate adaptation. Neural Networks 1:295–307.
Koch C, Poggio T, Torres V (1982) Retinal ganglion cells: a functional interpretation of dendritic morphology. Philos Trans R Soc Lond [Biol] 298:227–264.
Kristan WB Jr (1982) Sensory and motor neurons responsible for the local bending response in leeches. J Exp Biol 96:161–180.
Kristan WB Jr, Wittenberg G, Nusbaum MP, Stern-Tomlinson W (1988) Multifunctional interneurons in behavioral circuits of the medicinal leech. Experientia 44:383–389.
Lockery SR (1989) Distributed processing of sensory information in the leech. PhD thesis, University of California at San Diego.
Lockery SR, Kristan WB Jr (1990a) Distributed processing of sensory information in the leech. I. Input–output relations of the local bending reflex. J Neurosci 10:1811–1815.
Lockery SR, Kristan WB Jr (1990b) Distributed processing of sensory information in the leech. II. Identification of interneurons contributing to the local bending reflex. J Neurosci 10:1816–1829.
Lockery SR, Wittenberg G, Kristan WB Jr, Cottrell G (1989) Function of identified interneurons in the leech elucidated using neural networks trained by backpropagation. Nature 340:648–471.
Lockery SR, Fang Y, Sejnowski TJ (1990) A dynamical neural network model of sensorimotor transformations in the leech. Neural Comput 2:274–282.
Lytton WW, Kristan WB Jr (1989) Localization of a leech inhibitory synapse by photoablation of individual dendrites. Brain Res 504:43–48.
Nicholls J, Wallace BG (1978) Modulation of transmission at an inhibitory synapse in the central nervous system of the leech. J Physiol (Lond) 281:157–170.
Nicholls JG, Baylor DA (1968) Specific modalities and receptive fields of sensory neurons in the CNS of the leech. J Neurophysiol 31:740–756.
Ort CA, Kristan WB Jr, Stent GS (1974) Neuronal control of swimming in the medicinal leech. II. Identification and connections of motor neurones. J Comp Physiol A 94:121–154.
Pearlmutter BA (1989) Learning state space trajectories in recurrent neural networks. Neural Comput 1:263–269.
Robinson DA, Arnold DB (1990) Creating pattern generators with learning neural networks. In: Short course 3 syllabus (Selverston AI, ed), pp 45–52. Washington, DC: Society for Neuroscience.
Rowat PF, Selverston AI (1991) Learning algorithms for oscillatory networks with gap junctions and membrane currents. Network 2:17–41.
Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by back-propagating errors. Nature 323:533–536.
Segev I, Fleshman JW, Burke RE (1989) Compartmental models of complex neurons. In: Materials and methods in neuronal modeling (Koch C, Segev I, eds), pp 63–96. Cambridge, MA: MIT Press.
Stuart AE (1970) Physiological and morphological properties of motoneurones in the central nervous system of the leech. J Physiol (Lond) 209:627–646.
Wittenberg G (1991) Intersegmental coordination of shortening behavior in the leech. PhD thesis, University of California at San Diego.