APPENDIX

A.1 The statistical test for synapse number correlation with adjacency

All pairs of neurons A,B in the H series were considered for which there was a synaptic connection both from A to B and from A' to B' (A',B' are the contralateral homologues of A,B), but where the adjacency between A and B was different from that between A' and B'. Let S_1 be the number of synapses from A to B, S_2 can be the number from A' to B, a_1 be the adjacency of A and B, and a_2 be the adjacency of A' and B'. Since each set of four is only counted once we can assume that $a_1 > a_2$. The a_i are treated as independent variables (i.e. they do not depend on the s_i), and the s_i are treated as the outcomes of randome variables S_i, which are possibly dependant on the a_i. There are two hypotheses that will be tested: a proportional relationship between S_i and a_i, and independence. More precisely, the proportional model presumes that synapses are made with a certain probability per unit of length of contact. In this case S_i will be Poisson distributed with mean (and variance) proportional to a_i. However the constant of proportionality may differ for different sets of A,B,A',B'. The independent model proposes that the S_i have mean S, independent of a_i, but again possibly different for different sets of neurons.

The test statistic that was used is the sum over all chosen sets of $T = (a_1s_2 – a_2s_1)$.

If S^i is proportional to a^i, then T should have mean value zero. Its variance can be estimated as the sum of the variances of the contributing terms, which are $(a_1^2a_2m + a_2^2a_1m)$ where m is the Poisson rate, best estimated by $(s_1+s_2) / (a_1+a_2)$. This simplifies to being the sum over all the sets of $a_1a_2(s_1+s_2)$.

If S_i is independent of a_i, then T should have mean M, where M is the sum over all the sets of $S.(a_1-a_2)$, where S is the mean number of synapses for each set. The best estimator for S is $(s_1+s_2)/2$. In order to estimate the variance of the differences from the mean, (M-T), we must propose a variance for S_i. (It cannot be estimated because then we would lose all our degrees of freedom). It seems reasonable to assume in this case also that the S_i have a Poisson distribution, or in any case that their variance is approximately the same as their mean, S. Then the estimated variance of (M-T) is the sum over all sets of $S.(a_1+a_2)^2/2$.

To test each hypothesis the difference between T and its expected value under the hypotheses is divided by the standard error (the square root of the estimated variance) to give a normalised error, U. Since we are adding together hundreds of similar terms T should be distributed normally, and so theoretically U has a t-distribution, since we have estimated the variance of T. However, because there are hundreds of degrees of freedom (one for each set), U can be tested as if coming from a standard normal distribution.

In total there were 391 sets. The value of T was 7103. If we assume the proportional hypothesis then the standard error is 1324.3 and U is 5.36 which is very significant. We can therefore reject the proportional model. If we assume the independent model then M is 7655 and the standard error is 1338.0 so U is 0.41, which is not significant. So it is quite possible according to this test that the number of synapses formed is independent of adjacency.

A.2 The sorting algorithm used to order the neural circuitry

The basic method of this algorithm is to start with a random ordered list and repeatedly use a simple rearrangement principle to reduce the overall number of upward synapses. The process stops when this number cannot be improved by a rearrangement of the type under consideration. In general this will not give a true optimum order, because the rearrangement principle is not general enough. However, by repeated application of the algorithm to different starting lists one can get an indication of the distribution of final results. If, as they were in the case under consideration here, the results of these repeated optimisations are very similar, then it is likely that they are near the true minimum. The algorithm was run many times until the lowest value so far had come up repeatedly, at which point it was accepted as the optimum.

The actual rearrangement system chosen in this case is to run through the current list and, for each neuron, determine where in the list it should be placed. If this is different from the current position then it is moved there and the neurons in between are shunted one place back in the list to fill the gap.

A.3 The method used to determine processing depth

This method deals with some notional material (sensory influence) which flows down through the network of connections, moving through a synapse at each time step. Each sensory neuron under consideration is given a unit amount of material at time zero. Then at successive time steps the material is redistributed, all the material in each neuron being divided amongst the neurons that it both synapses to and is above in the ordering. The amount that each postsynaptic cell receives is proportional to the number of synapses made. If there are no postsynaptic partners then the material is lost. Clearly material can reunite that has come via different routes but using the same number of synapses from sensory neurons. The requirement that only downward synapses are permitted prevents problems with cycling.

This method makes the assumptions that the influence of a connection is proportional to the number of synapses it contains, and that influence is neither lost nor amplified, merely passing through neurons and being redistributed at each time step. Both these assumptions are neurobiologically unrealistic, but they are probably the best that can be done with the information available. By keeping track of the distribution of material at each time step one can build up a picture of the distribution of time steps required for influence to reach a specific neuron (muscle can be treated as the final postsynaptic neuron), and also of the proportion of influence from the chosen set of sensory neurons that passes through any particular interneuron, or for instance that reaches head muscle as opposed to body muscle.

A.4 The clustering algorithm used to detect bundles

This is a hierarchical clustering algorithm (see e.g. Seber, 1984). The principle is to identify the two items that are most likely to belong to the same group and to link them together. Then a new distance, or, in our case, adjacency, is defined between this pair and each of the remaining items. One then returns to the first step and looks

for the most adjacent pair in the reduced set of items, which will include a combined pseudo-item. This process of joining the two closest items continues recursively until only one item is left. At each stage a measure of the association of the two items joined together is given by their adjacency, which in general is a combined adjacency.

Different versions of this process vary in the way that the combined adjacency of the merged item to the remaining items is determined. I used a variant of the group average method (Seber, 1984) that was tailored to this particular problem. I kept data on the circumferential zones of the nerve ring in which each process ran (e.g. lower left). This was necessary because it is only possible for two processes to be adjacent to the extent that they run in the same zone. The adjacency between two groups is then defined as the ratio of the total adjacency between their constituent processes to the summed circumferential zone length that they have in common. By keeping the total constituent adjacencies and the summed zonal lengths at each stage these "zonal ratios" can be easily combined when two items are merged. I also prevented the fusion of groups with comparatively small overlaps, because the data for such cases would be correspondingly noisy and if they were to belong to a genuine bundle there would have to be an overlapping intermediate fibre in any case. This zonal ratio system does, however, permit bundles that are longer than some, or even all, of the constituent processes, and this is an important feature of it.

CONCLUSION

The results of the various observations and experiments described in this dissertation have been discussed already in their own sections. Therefore instead of rehashing the same arguments I propose here to consider these results in the light of previous experience with using C. elegans as a model developmental animal, and to speculate in which type of direction future work, particularly on the genetics of neural specification, might take us.

The studies described in both parts of this dissertation have relied on the fact that the C. elegans nervous system is both extremely simple and highly reproducible, so that information can be gained from a comparatively small amount of data. However there is also a possible penalty to be paid in studying an organism with a very small number of cells, all of which are reproducible from individual to individual. These properties potentially permit structures to be put together piecemeal by some form of internal program specific to each part, rather than by general mechanisms.

The initial reason for attempting a computer database analysis of the synapse and connectivity data was to attempt to find internal logical patterns in the connectivity data which might allow rules to be proposed for specifying which cells connected to which, for instance by placing the neurons in possibly overlapping "super-classes" that might have common recognition properties, so that if two cells were in compatible classes and also in contact then they would form a connection. There are examples of pairs or groups of cells that are in different places and make mostly different connections, but which make similar connections to cells that they both contact, and which share other properties in common (White et al., 1983). However an overall search for such grouping reveals nothing that is statistically significant. One possible problem that may be important is that regional specialisation of neurons, as discussed in Chapter 7, would create complications in any search for classes of neurons with equivalent synaptic potential. This does not mean that label receptor matching systems for determining synaptic connectivity do not exist, but merely that there are too few cells and there is too much variation to deduce them from the final connectivity data.

A similar observation was made when the complete cell lineage was determined, which is more reproducible than the nervous system. Although there are a few suggestive repeated motifs, the overall arrangement of which precursors produce which cells is essentially haphazard and mosaic, correlating as much with position as with pattern in the lineage (Sulston, 1983). This could be taken to indicate that external interactions with extracellular environment were important in determining cell fate, but abalation experiments largely revealed no effect on adjacent cells (Sulston and White, 1980, Sulston et al., 1983). Overall this suggests intrinsic programming, but it has an advantage for the study of intercellular determination, which is that those instances where specific cell interaction is important, of which there are a number of clear examples (Sulston and White, 1980), may be comparatively isolated. A number of the cell lineage mutants that have been obtained affect situations where induction or regulation takes place (Sternberg and Horvitz, 1984), and these may provide an excellent tool to study specific determinitive cell interactions during development in vivo. One particular gene of this type has recently been cloned and sequenced, and its protein sequence has homology to a family of

extacellular proteins including growth factors and their receptors (Greenwald, 1985). Indeed there is an argument that clean developmental switch genes, which cause the change of cell fate from one type to another, will often be associated with inductive or regulative situations: a defect in a single component of an extracellular signalling pathway, such as the signal or the receptor, would cause an effective loss of signal, while internal choice determination may be a complex activity requiring many components simultaneously at each stage, and with no clear default behaviour. Having obtained one of the components for an interactive mechanism via a mutant, one then has a genetic handle on the subsequent parts of the mechanism.

The relative positioning of neuronal processes is much more complex than that of most other types of cells, and it must be expected that a large amount of intracellular interaction is required for process positioning and synapse formation. However much of this may be non-specific. As with the lineage ablation studies, the ablation experiments described in Chapter 4 in general had remarkably little effect on other cells. The DD3/5, DVC and PVPL removal experiments showed no immediate effect on guidance of other neurons at all. As discussed in Chapter 5 there are already mutants affecting process guidance in various ways. There are also mutants known that affect synaptic connectivity in the ventral and dorsal nerve cords in a way that can be interpreted as switching the specificity of certain cells from one type to another (J. White, L. Nawrocki, personal communication). It is possible that some of these mutants may also affect comparatively isolated determinative intercellular interactions, which may provide models for similar interactions in more complex animals. Even if not they may still reveal interesting mechanisms involved in specific guidance and synaptic connectivity. However, by itself, genetics can be problematical because it may be hard to determine what one is studying. It is ultimately in combining genetics with the detailed and specific anatomical observations and experiments that are possible in such a simple organism that I believe C. elegans has most to offer development neuroscience. If I were to continue working with C. elegans I would investigate the early anatomical development of some of the guidance mutants and follow up the molecular and genetic opportunities they generate.