THE ROYAL SOCIETY

doi 10.1098/rstb.2001.0905

Generation, description and storage of dendritic morphology data

Giorgio A. Ascoli[1,2*], Jeffrey L. Krichmar[1,3], Slawomir J. Nasuto[1] and Stephen L. Senft[1]

[1]*Krasnow Institute for Advanced Study, George Mason University, MS2A1-4400 University Drive, Fairfax, VA 22030-4444, USA*
[2]*Department of Psychology, George Mason University, MS2A1-4400 University Drive, Fairfax, VA 22030-4444, USA*
[3]*The Neurosciences Institute, 10640 J. J. Hopkins Drive, San Diego, CA 92121, USA*

It is generally assumed that the variability of neuronal morphology has an important effect on both the connectivity and the activity of the nervous system, but this effect has not been thoroughly investigated. Neuroanatomical archives represent a crucial tool to explore structure–function relationships in the brain. We are developing computational tools to describe, generate, store and render large sets of three-dimensional neuronal structures in a format that is compact, quantitative, accurate and readily accessible to the neuroscientist. Single-cell neuroanatomy can be characterized quantitatively at several levels. In computer-aided neuronal tracing files, a dendritic tree is described as a series of cylinders, each represented by diameter, spatial coordinates and the connectivity to other cylinders in the tree. This 'Cartesian' description constitutes a completely accurate mapping of dendritic morphology but it bears little intuitive information for the neuroscientist. In contrast, a classical neuroanatomical analysis characterizes neuronal dendrites on the basis of the statistical distributions of morphological parameters, e.g. maximum branching order or bifurcation asymmetry. This description is intuitively more accessible, but it only yields information on the collective anatomy of a group of dendrites, i.e. it is not complete enough to provide a precise 'blueprint' of the original data. We are adopting a third, intermediate level of description, which consists of the algorithmic generation of neuronal structures within a certain morphological class based on a set of 'fundamental', measured parameters. This description is as intuitive as a classical neuroanatomical analysis (parameters have an intuitive interpretation), and as complete as a Cartesian file (the algorithms generate and display complete neurons). The advantages of the algorithmic description of neuronal structure are immense. If an algorithm can measure the values of a handful of parameters from an experimental database and generate virtual neurons whose anatomy is statistically indistinguishable from that of their real counterparts, a great deal of data compression and amplification can be achieved. Data compression results from the quantitative and complete description of thousands of neurons with a handful of statistical distributions of parameters. Data amplification is possible because, from a set of experimental neurons, many more virtual analogues can be generated. This approach could allow one, in principle, to create and store a neuroanatomical database containing data for an entire human brain in a personal computer. We are using two programs, L-NEURON and ARBORVITAE, to investigate systematically the potential of several different algorithms for the generation of virtual neurons. Using these programs, we have generated anatomically plausible virtual neurons for several morphological classes, including guinea pig cerebellar Purkinje cells and cat spinal cord motor neurons. These virtual neurons are stored in an online electronic archive of dendritic morphology. This process highlights the potential and the limitations of the 'computational neuroanatomy' strategy for neuroscience databases.

Keywords: ARBORVITAE; computational neuroanatomy; dendritic morphology; L-NEURON; virtual neurons

1. INTRODUCTION

Neurons constitute the structural building blocks for the brain, yet their own structure is far from elementary. Dendritic and axonal trees stemming from the soma develop an extremely complex pattern of branching that occupies and fills three-dimensional (3D) space. Since the original work of Santiago Ramón y Cajal, it has been apparent that the detailed knowledge of neuritic structures was an important aspect of neuroscience (Cajal 1894–1904). More recently, discoveries that many dendrites conduct input signals actively, back-propagate action potentials and integrate synaptic inputs by means of time-dependent nonlinear summation (reviewed e.g. in Eilers & Konnerth 1997) provided indisputable evidence that dendritic morphology is a key aspect of the neuronal machinery underlying signal processing and integration.

*Author for correspondence (ascoli@gmu.edu).

Equally important lines of research proved that network connectivity is constantly reshaped by the dynamic remodelling of both dendrites and axons, and that dendritic morphology is pivotal in determining the pattern of synaptic formation among neurons (recently reviewed in Wong & Wong 2000).

In order to explore the structure–function relationship in single neurons, electronic databases of complete and accurate 3D dendrites constitute a valuable tool. For example, neuroanatomical databases can be used to construct detailed electrophysiological models of ionic conductance (e.g. De Schutter 1999), study the effect of dendritic morphology on firing patterns (Krichmar *et al.* 2001), define anatomical classes unambiguously (Cannon *et al.* 1999), test hypotheses of growth mechanisms (Van Pelt *et al.* 1999) and investigate the influence of dendritic morphology on axonal growth (Van Ooyen *et al.* 2000). To date, few neuronal databases are publicly available (e.g. Cannon *et al.* 1998), but additional morphological data is made available by single research groups in sparse fashion and upon request.

Two crucial issues in the development and use of neuromorphological databases concern the experimental data acquisition and the format of the entries. Data acquisition refers to a multi-step process from tissue staining to the extraction of single neuron structural information. This is a difficult task that limits the number of entries stored in a database. The morphological data format refers to the code adopted to store neuromorphological data as entries into a database. There is a trade off between the accuracy guaranteed by digital format and the greater insight ensured by the usage of statistical distributions. The next two sections of this introduction briefly review the current scientific state of these important issues, with an emphasis on the unsolved problems. The remainder of this paper discusses the use of computational algorithms as an alternative approach to obtaining and describing neuroanatomical data.

(a) *Issues in data acquisition*

The experimental acquisition of single neuron morphological data is an arduous procedure. Neuronal processes normally are so intricately intermingled that for most species one must find means for selectively labelling individual cells in order to evaluate their structure. Nowadays this can be achieved in a great variety of ways, such as: superfusion with vital dyes (Muller 1989); impregnation with heavy metals (e.g. Cohen *et al.* 1973; Somogyi 1990); targeted iontophoresis or bulk injection with chemical chromophores (Stewart 1981; Purves & Hadley 1985; Godement *et al.* 1987) or with dye-amplifying enzymes (Muller & McMahan 1976); infection with gene sequences, either acutely (Luskin *et al.* 1988; Price *et al.* 1987) or heritably (see Stearns 1995) and even to developmentally selectable targets (Roy *et al.* 2000).

Once labelled, the task to examine and acquire a neuron's morphological features remains non-trivial. For more than a century, the principal tool for recording neuronal structures has been the light microscope. Its practical resolution has improved gradually from around 0.5 μm to better than 0.1 μm (Gustafsson *et al.* 1999). This degree of precision has long been sufficient to reveal the hallmark features of neurons, i.e. discrete somata sprouting a wide range of closely apposing dendritic and axonal branching processes. For many decades now, these features of central nervous systems have been routinely amenable to careful visual inspection, and have been documented in part by manual drawings and by photomicrographs.

However, without an unequivocal measurement standard, all labelling methodologies contend with the questions of whether neurons are labelled completely and whether any important structural changes were imposed by the method of tissue preparation. The resolution of the light microscope is not enough to settle all such questions reliably. For instance, it cannot always disambiguate fibre crossing from branching (especially because of poor depth resolution, discussed below). In addition, when more than one neuron is labelled, it is often impractical to distinguish which cell adjacent processes belong to, or whether neighbouring processes actually make contact (e.g. Cabirol-Pol *et al.* 2000).

These ambiguities in deciphering arbour structure are greatly exacerbated by physical sectioning and fixation. Sectioning (followed by serial reconstruction) is necessitated by the diminishing depth-of-field of the high-magnification objectives needed to resolve details. Fixation limits tissue autolysis, thereby preserving specimens for long-term observation. Both procedures compromise the ability to understand network organization by distorting morphological features, e.g. by artefactually affecting neurite diameter, inducing varicosities and creating discontinuities.

In order to overcome some of these limitations, one can resort to electron microscopy for resolving fine anatomical structures down to the resolution of single synapses. This technique is susceptible to additional interpretative confusions due to fixation. The major problem of electron microscopy, however, is the intensive manual labour involved in obtaining and manipulating hundreds or thousands of extremely delicate sections. With some notable exceptions (Stevens *et al.* 1980; Sims & Macagno 1985; Schierenberg *et al.* 1986; White *et al.* 1993; Soto *et al.* 1994; Shepherd & Harris 1998), these problems make electron microscopy an impractical method for the systematic extraction of neuronal structure, particularly over large areas. Nonetheless, electron microscopy plays an essential role in demonstrating the limits to accuracy and completeness of neuronal labelling.

A common problem of both light and electron microscopy is constituted by the z-resolution limit, which is usually much cruder than the resolution in the plane ('xy'). Widespread research efforts have yielded significant improvement of the depth resolution for light microscopy. For instance, numerous interference methods (see Inoue 1988) convert small differences of path length (due to changes in refractive index) into contrast, via constructive and destructive processes. More recently, improved discrimination in depth has come from advances in imaging with polarized light (Oldenbourg & Mei 1995) and from the confocal microscope (see Pawley 1990), particularly in its many recent multi-photon configurations (see Piston 1999). These tools, with the addition of computer-controlled fine focus, generate volumetric data in the form of stacks of aligned serial sections digitized at 0.5 to 0.1 μm increments along the z-axis. Such virtual (optical)

sectioning methods dramatically reduce the number of cutting artefacts otherwise generated by physical sectioning at 1 to 50 μm. In addition, the enhanced contrast captured by electronic scanning can also yield useful improvement of lateral resolution (see Oldenbourg *et al*. 1993). A final increase in overall resolution is obtainable from mathematical deconvolution of the resulting volume (see Rizzuto *et al*. 1998; McNally *et al*. 1999).

A further important source of uncertainty in the experimental acquisition of morphological data is constituted by the extraction of neuronal structures from microscopic images, however they are generated. This process, commonly referred to as 'tracing', is classically carried out with the aid of *camera lucida*, a system of lenses allowing the overlay of the image of 'external' object (e.g. paper and pencil) onto the microscope field of view. This task of manual labour can be partially computerized by using motorized stages controlled by the movement of a mouse or a joystick, interfaced by specialized software such as NEUROLUCIDA (Microbrightfield, Colchester, VT) (Glaser & Glaser 1990). Even in the computerized version, however, neuronal tracing involves a great deal of user interaction, and is plagued by low precision in the determination of the exact focus, which limits the effective z-resolution.

Recent attempts to automate this process completely are based on 'flood-filling' algorithms operating on volumetric data: starting from an arbitrary point within the scanned neuron, adjacent regions in the volumetric scan are automatically recruited if they 'belong' to the cell (Senft 1995). When the image has a sufficiently high contrast, the test to determine whether a voxel belongs to the structure being traced is based on a simple threshold criterion, although more sophisticated methods (Cohen *et al*. 1994) also can be used. The resulting connected object can be further processed by 'thinning' (e.g. Gong & Bertrand 1990) in 3D, which provides the complete list of orientated lengths, bifurcations and terminations in the branching structure. A pre-processing step allows one to preserve information on the original diameter (Senft 1995).

This automated computer-based procedure allows one to obtain highly detailed neuronal structures in a much shorter time than with manual tracing (minutes rather than days), provided that the contrast of the image is adequate. In addition, the assignments of neurite diameter and z-values are more reproducible than with hand tracing. Remaining unresolved ambiguities, such as crossing points appearing as chiasms or loops, can be edited out manually or semi-automatically based on heuristics. These algorithms sometimes amplify surface irregularities and mistakenly interpret visual noise as cell structures. These problems, too, can be resolved with manual editing or judicious smoothing.

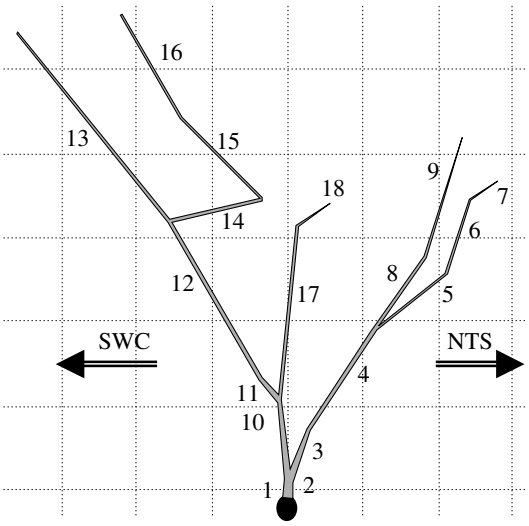
The major drawback of algorithmic neuronal reconstruction is that the automatically traced data is limited to the scanned volume. The recent availability of computational memory (RAM) in the gigabyte range makes it possible to analyse several adjacent fields at once. For relatively compact trees, this allows the automatic tracing of entire trees. However, this possibility is less suitable to apply to large dendritic fields (as in spinal cord motor neurons), and to axons, often meandering well beyond the confines of the lens field of view, and even beyond the relatively thick physical sections cut for confocal microscopy.

(b) *Morphological data formats*

Raw data acquired experimentally can be stored in different formats. The most classical possibility is to represent the dendrite with a picture or set of pictures. Although this is the most commonly used format (e.g. in publications), and yields intuitively accessible information, pictures are flawed by the major problem of data loss. A 3D structure is represented two-dimensionally, and even if multiple perspectives can be used, the retrieval of information depending on the third dimension (in any perspective) is rarely complete. An obvious alternative is to represent the 3D structure of dendrites in a digital format, allowing 'pseudo-3D' rendering and animation in modern computer graphics. In this format, the branches of a dendritic tree are represented as a set of cylinders, and each cylinder is described by a line of the morphological file. The three main digital formats adopted to describe dendritic morphology are the 'SWC' format (Cannon *et al*. 1998), the 'Eutectic' format (Capowsky & Schneider 1985 and references therein), and the 'Neurolucida' format (Glaser & Glaser 1990).

In general, the digital formats consist of a plain text file with each line describing the geometry of one neuronal segment (cylinder). These files can range from 500 lines for small, simple cells, to 10 000 or more lines for large and complex cells. In the SWC format, dendritic segments are characterized by an identification number, a type (to distinguish basal, apical, proximal, distal and lateral trees), the x, y, z positions of the cylinder ending point (in μm with respect to a fixed reference point), a radius value (also in μm), and the identification number of the 'parent', i.e. the adjacent cylinder in the path to the soma (the parent of the root being the soma itself). In the Eutectic format, dendritic segments are characterized by a type, the Cartesian positions of the cylinder ending point, a diameter value, and the topological assignment of the cylinder, i.e. whether the segment is continuing (it has one 'daughter'), bifurcating (two daughters), or terminating (no daughters). The SWC and Eutectic formats are thus very similar, and it is easy to verify that both conserve all the information available in the dendritic morphology. Some research groups use variations of these two formats, and simple software routines can be written to interconvert them. An example of these two formats is reported in figure 1. Interestingly, the new Neurolucida format (version 4.0) adopts a 'pseudographic' convention, in which each segment is characterized by a type, Cartesian position and diameter, while the connectivity is determined by the number of indentations in the line (at each bifurcation, daughters are indented with respect to their parent, and terminations are marked by a horizontal line). The digital description of the SWC, Eutectic, or Neurolucida format constitutes a completely accurate mapping of dendritic morphology, but it bears little intuitive information for the neuroscientist (e.g. it is difficult to establish the morphological class of a neuron by simply looking at its SWC or Eutectic file). In addition, the Cartesian format is not compact (a typical single neuron is described by several thousand lines).

ID	T	X	Y	Z	R	C
1	1	−1.5	−8.2	1.3	3.8	−1
2	3	0.8	4.2	1.4	2.3	1
3	3	21.1	41.1	1.4	1.3	2
4	3	89.2	115.1	−5.3	1.1	3
5	3	153.6	163.5	3.0	0.6	4
6	3	184.9	218.1	−2.3	0.5	5
7	3	220.9	238.7	−13.7	0.3	6
8	3	159.9	183.9	−23.6	1.1	4
9	3	211.9	270.9	−35.2	0.7	8
10	3	0.6	58.1	−2.6	1.1	2
11	3	−0.5	75.9	−18.6	1.1	10
12	3	−40.5	180.0	−55.3	1.0	11
13	3	−156.4	281.8	−65.9	0.8	12
14	3	0.3	186.8	−7.0	0.8	12
15	3	−51.5	235.7	−27.0	0.8	14
16	3	−85.2	297.6	−41.5	0.5	15
17	3	75.3	198.3	−54.6	0.8	10
18	3	130.6	228.6	−81.3	0.5	17



T	X	Y	Z	D	C
1	−1.5	−8.2	1.3	7.6	s
3	0.8	4.2	1.4	4.7	b
3	21.1	41.1	1.4	2.7	c
3	89.2	115.1	−5.3	2.2	b
3	153.6	163.5	3.0	1.3	c
3	184.9	218.1	−2.3	0.9	c
3	220.9	238.7	−13.7	0.6	t
3	159.9	183.9	−23.6	2.2	c
3	211.9	270.9	−35.2	1.4	t
3	0.6	58.1	−2.6	2.2	b
3	−0.5	75.9	−18.6	2.2	c
3	−40.5	180.0	−55.3	2.0	b
3	−156.4	281.8	−65.9	1.5	t
3	0.3	186.8	−7.0	1.5	c
3	−51.5	235.7	−27.0	1.5	c
3	−85.2	297.6	−41.5	1.0	t
3	75.3	198.3	−54.6	1.5	c
3	130.6	228.6	−81.3	1.0	t

Figure 1. Digital representation of dendritic morphology. This extremely simplified neuronal structure was obtained by extensive pruning of a dentate gyrus granule cell (Cannon *et al.* 1998). The left panel shows the digital SWC representation: every cylindrical compartment is described by a row containing a label (ID, the same numbers reported next to the branches in the picture, middle panel), tag (T, 1 for soma, 3 for dendrite), Cartesian positions (x, y, z, in micrometres), radius (R), and connectivity (C, representing the label of the parent; −1 indicates no parent). The right panel shows the NTS digital representation of the same structure (the acronym stands for Neuronal Tracing System, the original name of Eutectic). The only differences are the absence of a label, the choice of diameter (D) to represent the thickness of the cylinder, and the connectivity format that refers to the cylinder topology: 's' is the soma, 'b' a bifurcating segment (two daughters), 'c' a continuing segment (one daughter), 't' a terminating segment (no daughters).

In contrast to the digital format, a classical neuroanatomical analysis characterizes neuronal dendrites on the basis of statistical distributions of morphological parameters derived from the raw data (Uylings *et al.* 1986). A typical example is the widely adopted Sholl analysis (Sholl 1953), which plots the number of dendritic segments as a function of the distance from the soma. Other commonly used morphological descriptors are branch length statistics, such as length distributions versus size of soma, versus number of branches, versus branch diameter, or versus position in the layer (Ishizuka *et al.* 1995), tree shape, such as depth over width ratio (Claiborne *et al.* 1990), bifurcation 'partition' or asymmetry (Van Pelt *et al.* 1992), or any correlation between the above parameters (Larkman 1991). This description is intuitively more accessible, but it only yields information on the collective anatomy of a group of dendrites, i.e. it is not complete enough to provide a precise 'blueprint' of the original data. In addition, from classical neuroanatomical analyses, it is usually impossible to extract more morphological information than that which is explicitly reported. For example, if someone were interested in knowing the distribution of the terminal diameter versus their dendritic path distance from the stem, this information would not be available from other parameter distributions and correlations.

In summary, dendritic morphology databases in Cartesian format (SWC, Eutectic, Neurolucida, or similar formats) contain complete information that can be used to derive any morphological measurement, but they require extensive storage space and are not intuitively accessible from the geometrical or physical point of view. In contrast, dendritic morphology databases based on pictures or on classical statistical plots are compact and intuitively accessible to the neuroscientist, but they do not conserve the entirety of the information contained in the original data.

(c) *The 'computational' alternative*

The computational neuroanatomy approach to dendritic morphology databases consists of the algorithmic generation of neuronal structures within a certain morphological class, based on a set of measured 'fundamental' parameters. This strategy has great potential to overcome the problems related to both extensive experimental acquisition and format of single-neuron morphological data. Given the right type of algorithm, this level of description is as intuitive as a classical neuroanatomical analysis (because the statistical distributions of the fundamental parameters have an intuitive geometrical meaning), and as complete as the digital format (because these parameters are sufficient to generate and display complete neurons). Since the fundamental parameters measured from experimental data result in statistical distributions, algorithms that generate 'virtual neurons' sample values from these distributions stochastically. As a result, just as in nature, no two virtual neurons are identical, even if they belong to a recognizable anatomical class (Ascoli 1999).

This 'computational' approach to single-cell neuroanatomy, originally proposed in the late 1970s (Hillman 1979), can now be fully exploited thanks to the recent explosion in computational power and computer graphics tools. With only a limited number of experimentally traced neurons for each morphological class, a large number of 'virtual' database entries can be generated, thus minimizing one component of the time-consuming experimental process. At the same time, an entire

morphological class can be completely described by a handful of parameters, thus offering an ideal solution for the morphological database formats. The next section of this paper describes the details of the major computational neuroanatomical algorithms for single neurons. The following two sections describe an example of the application of this strategy to an electronic database of virtual morphology and the rationale for its validation. Finally, the last section of the paper discusses advantages and disadvantages of individual algorithms as well as of the general computational neuroanatomy approach.

2. DESCRIPTION OF THE ALGORITHMS

Two major types of computational algorithms have been proposed for the virtual generation and description of dendritic trees, namely 'local' and 'global' algorithms. Local algorithms rely entirely on a set of local rules intercorrelating morphological parameters (such as branch diameter and length) to let each branch grow independently of the other dendrites in the tree as well as of its absolute position within the tree. In global algorithms, new dendritic branches are dealt 'from outside' to competing groups of growing segments, also depending on their position in the tree (e.g. on their distance from the soma).

Local and global algorithms offer complementary advantages. Local algorithms are simpler, are more intuitive, and the fundamental parameters used by the algorithms can be measured directly from experimental data. Because of their small number of parameters, they are well suited to study structure–function relationships and the origin of emergent properties (i.e. anatomical parameters that are not explicitly imposed in the algorithm). Global algorithms are usually more flexible, but many of their fundamental parameters must be obtained through extensive and elaborate parameter searches. Global algorithms can also be extended to generate populations of interconnected neurons (networks), instead of single neurons. In the next three sections, we describe a series of local and global algorithms as implemented in two software packages, L-NEURON (Ascoli & Krichmar 2000) and ARBORVITAE (Senft 1997; Senft & Ascoli 1999), as well as other relevant neuroanatomical algorithms and rules.

(a) *Local algorithms in L-NEURON*

L-NEURON is a program designed to describe, generate and visualize the precise dendritic morphology of single neurons. The L-NEURON executables (for Unix and Windows) are publicly available at http://www.krasnow.gmu.edu/L-Neuron/index.html (case sensitive). The software implementation is continuously improved and additional interfaces, conversion algorithms and measurement tools are under development. L-NEURON is based on a modification of an original program to generate L-systems architectures, including fractals and botanical trees (Prusinkiewicz & Lindenmayer 1990). As such, it can read in regular L-system files (in the 'turtle graphics' dialect adopted by Laurens Lapré, http://www.xs4all.nl/~ljlapre/lparser.htm) and produce any type of geometrical recursive structure. In addition to the generation of L-systems, L-NEURON implements a series of local neuroanatomical rules to create virtual neurons in 3D space (Ascoli & Krichmar 2000). The intermediate output of both the L-systems and the neuroanatomical algorithms is a 'production string', i.e. a sequence of characters with specific geometrical meanings in terms of 'drawing' commands (such as grow forward, split, turn by an angle, taper, etc.). The production string can be then translated into digital format in a variety of outputs, including SWC, VRML, POV, DXF, or VOL, a binary format that can be displayed by a viewer under Windows. Thus, L-NEURON products can be both represented graphically and used in electrophysiological simulations.

The first class of neuroanatomical rules implemented in L-NEURON was derived from an original proposal by Hillman (1979, 1988). According to Hillman, a dendritic tree can be described (and generated) as a sequence of local, recursive processes corresponding to the growth of single branches starting from a stem or a bifurcation point, and ending with another bifurcation point or a terminal tip. A branch (characterized by a certain starting diameter) grows for a certain length and tapers. If its final diameter is larger than a threshold, it bifurcates (giving rise to two new branches), otherwise, it elongates for an additional terminal amount and it stops. Upon a bifurcating event, the diameters of the two daughters are determined from the final diameter of the parent via the ratio between the two daughter diameters and the generalized Rall's 'power rule', stating that the sum of daughter diameters elevated to a coefficient equals the parent's diameter elevated to the same coefficient (see Hillman 1979). In Hillman's description, a branch stops growing not because of a fixed number of iterations (as in L-systems), but because all branches, through tapering and bifurcations, end up with too small a diameter to grow further.

Hillman called the branch length, taper rate, diameter threshold for bifurcations, terminal length, daughter diameter ratio and Rall's power coefficient, 'fundamental parameters of shape' (Hillman 1979), viewing them as a complete set of descriptors for the dendritic tree. In L-NEURON, Hillman's 'algorithm' is implemented with the addition of two elements that make the process truly complete. First, angles are introduced to orientate dendritic stems in three dimensions (by an azimuth and an elevation with respect to the centre of the soma), and to describe or generate bifurcations (by an amplitude and a twist of the plane). This simple characterization of bifurcation angles assumes that bifurcations have a uniformly random tilt, and that the parent and the two daughters lie on the same plane. Both of these assumptions are only very rough approximations of reality for all the neuronal classes tested (S. L. Senft and G. A. Ascoli, unpublished data). Second, Hillman's algorithm neglects to describe the dendritic path within a branch, which has a clear influence on important characteristics such as dendritic length and overall tree size. In L-NEURON, the branch path is characterized by a 'fragmentation' process in which a branch is described as a sequence of shorter segments. The fragmentation 'smoothness' (i.e. amplitude of connection angles between consecutive segments) is determined by the measured ratio between the branch path length and the Euclidian distance between branch starting and ending points.

All Hillman's fundamental parameters of shape, as well as angles and fragmentation in L-NEURON, can be

measured experimentally from a set of neurons belonging to the same morphological class. For example, bifurcation amplitude angles can be measured at each bifurcating point in each available dendrite of a given type from the neuronal set. Obviously, these measurements are easier, faster and more reliable if the morphological data are available in digital format. Each of the resulting sets of L-NEURON parameters can conform to a variety of statistical distributions, such as Gaussian, uniform, constant, gamma, exponential, etc. The set of statistical distributions describe the morphological class from which they were extracted. When L-NEURON is run to generate neurons, it samples parameter values stochastically within the correct statistical distributions. Every virtual neuron grown in such a process will be a unique individual (i.e. as in nature, no two virtual neurons are identical). However, if a set of virtual neurons is produced, their collective morphological properties (as measured by the fundamental parameters) will be statistically identical to those measured from the real neurons.

Two additional variations have been implemented in the Hillman-like algorithm in L-NEURON. First, Rall's power equation was modified in a less restrictive way by multiplying the parent diameter by a constant that we called 'poliko' (power linear 'konstant'), to respect experimental findings more faithfully (e.g. Cullheim *et al.* 1987; Hillman 1988). Second, an option is given to calculate the bifurcation amplitude and tilt angles with Tamori's equations (Tamori 1993), instead of measuring the amplitude and sampling a random tilt. Tamori's 'effective volume' analysis logically connects bifurcation angles with the parent diameter and with a parameter (the effective dimension) that is constrained by Rall's power as an upper bound. In L-NEURON, Tamori's variation to Hillman's algorithm is implemented by sampling a random effective dimension between the value of one and the sampled value for Rall's power, and using this number to calculate tilt and amplitude (Tamori 1993; Ascoli & Krichmar 2000).

The second, alternative class of neuroanatomical rules implemented in L-NEURON corresponds to the simplest of the models proposed by Burke and co-workers to describe spinal cord motor neurons (Burke *et al.* 1992). This algorithm is also local, stochastic and recursive, in that the growth of each dendritic segment is only determined by its diameter and random sampling and, upon growth, a segment spawns more segments that obey the same rules. Burke's model describes dendritic growth also within a branch, i.e. it allows 'extension' processes on top of bifurcations and terminations. At each step of the algorithm, branches grow by a unitary length and taper their diameter. A random number is then sampled and tested to check whether it falls below a diameter-dependent value of a bifurcation probability distribution. If this test is passed, the branch bifurcates and generates two daughters. Otherwise, a second analogous sampling occurs with a termination probability distribution. If the termination test is passed, the branch stops—otherwise it starts a new iteration with unitary growth, etc. Burke's model differs from Hillman's also for the determination of the daughter's diameters at bifurcations: instead of using the diameter ratio and (a modified) Rall's law, daughter diameters are sampled empirically from a statistical distribution derived from the experimental correlations of the normalized diameter pairs (i.e. the measured linear correlation in the scatter plot of the diameter of one daughter divided by the parent diameter versus the diameter of the other daughter, divided by the parent diameter). Like Hillman's, Burke's algorithm does not determine the value of the angles, which are thus implemented in L-NEURON exactly in the same fashion.

Despite being formulated specifically for motor neurons, Burke's model is more flexible than Hillman's. However, it uses parameters that do not have a direct biophysical meaning (such as the probability of bifurcation or the sampling parameter for daughter diameters). In contrast, most of Hillman's parameters correspond to a precise subcellular rationale. For example, Hillman's minimum threshold diameter for bifurcation reflects the impossibility of a branch to split when only containing enough microtubules to support one process (Hillman 1979, 1988). In other words, Burke's model seems to describe dendritic morphology more accurately at the empirical level but less satisfactorily at the level of underlying biological mechanism. Burke's description can be improved by adding semi-global and global constraints to the local algorithm. Particularly, the distribution of daughter diameters appear to be affected by the diameter of the grandparent (Burke *et al.* 1992), while the bifurcation and termination probability distributions depend not only on the diameter but also on the path distance from the soma (Nowakowski *et al.* 1992).

At present, L-NEURON implements a single global parameter that seems to influence dendritic shape dramatically in all implemented algorithms, namely 'tropism' (Prusinkiewicz & Lindenmayer 1990). Tropism is a post-processing modification of the bifurcation angles (i.e. of the directions of branches), which 'pushes' cylinders in one of the Cartesian directions or away from (or towards) the soma. While tropism was originally used in L-systems to mimic the effect of gravity and wind, in dendritic growth it seems suitable to model the presence of gradients of trophic factors.

(b) *Global algorithms in ARBORVITAE*

ARBORVITAE is a program motivated by a desire to visualize activity patterns in large networks of interconnected and developing neurons (Senft 1997). To a certain degree it is possible to obtain such information experimentally through optical imaging (Orbach *et al.* 1985; Haglund *et al.* 1992; Dailey & Smith 1994; Tsodyks *et al.* 1999; Shoham *et al.* 1999). When dealing with an entire nervous system, however, optical imaging is limited by physical constraints in depth-of-field, field-of-view, resolution and signal-to-noise with depth, and by a restricted ability to discriminate blended signals arising from disparate physiological processes. In order to bypass these problems, ARBORVITAE aims at synthesizing brain anatomy and physiology, from the subcellular toward the system level, using computational simulations. An intention of ARBORVITAE is to combine information from the experimentally available vignettes of nervous system structure and activity to emulate the genesis, outgrowth and interactions among sets of neurons. The core representation constitutes skeletal 3D branching structures of neurons described statistically at the level of populations.

Thus, one of the aspects simulated by ARBORVITAE, is dendritic morphology. Because of its more ambitious goals, ARBORVITAE is more complex than L-NEURON, despite many common conceptual aspects. At the present time, ARBORVITAE is not publicly distributed to the neuroscience community.

ARBORVITAE instantiates individual neurons as members of groups. Morphological properties of each neuron are specified by sampling a set of random variables defined at the group level. Every variable is governed by statistical distributions fitted to experimental data. Basic examples include the locations, sizes and orientations of somata, or the orientation angles, length and diameter of dendritic segments. Neurons emit primary neurites that emerge from the soma in number, location and direction characteristic for the group. In general, although some growth decisions in ARBORVITAE are made using also local information, many constraints are predominantly imposed at the group level. This global approach constitutes the most striking difference from the local algorithms implemented in L-NEURON.

A biological rationale behind global algorithms lies in the hypothesis that arborization in neurons is resource-limited (see Samuels *et al.* 1996): a soma of a given size might trophically support only a certain length of distal neurite, and in a mature neuron, expansion of one subtree can promote the regression of another. ARBORVITAE extends this concept to the level of the cell group: a certain amount of growth resource is available to a group of cells as a whole (and it is therefore determined in the program at the group level). The growth resource is then dealt out progressively to all cells, in growth epochs. This has the effect of imposing approximately similar growth on all of the cells of a group in each epoch. Epochs represent different stages of growth in the development of the neuron. Between growth epochs, the statistical distributions of morphological parameters may vary (as determined at the group level). The precise growth pattern of each dendrite in a group is negotiated with all other 'active' sites (i.e. portions of the growing dendrites that could receive additional segments) that compete for the remaining group resource quota.

Arbours are constructed as linked tubular segments that can meander, branch and taper, and that can be detailed with synapses and varicosities or spines. Statistical distinction is made between segments that 'append' to somata (thus initiating neurites) and those that 'extend' or 'bifurcate' (thus continuing neurites). Different morphological classes result from different statistical distributions of morphological parameters, as well as from varying the proportion of appending, extending and bifurcating branches. Finally, a 'growth failure' probability is also specified at developing branches to enhance topological and geometrical variability. Within this general framework, ARBORVITAE implements two alternative global algorithms of growth.

In the first algorithm (Senft 1997; Senft & Ascoli 1999), dendritic segments are first dealt out to somata in 'appending' mode, to initiate neurites. Dealing occurs in round-robin fashion with a 'skipping' probability. The resulting distribution of dendritic stems per soma is Gaussian, with mean and standard deviation controlled by the number of appending segments and the skipping probability (G. A. Ascoli and S. L. Senft, unpublished data). The orientation of the appending segments is governed independently by additional random variables specified for the group. Next, additional dendritic segments are dealt progressively to pre-existing segments, again in numbered order and with a failure probability. A certain percentage of the quota of dendritic segments is dealt in 'extending' mode (i.e. only terminal tips can grow). Then a second portion is dealt in 'bifurcating' mode (i.e. 'internal' segments receive a second daughter and terminal segments receive a pair of daughters). The remaining quota is used for terminal extension. Since all growing sites in any epoch are labelled with a numerical identity, this dealing process also allows the user to specify a 'starting point' within the set of active sites below which no growth is permitted, by imposing the constraint of a minimum identity number threshold for a site to receive a segment. This additional parameter can permit only a subset of cells to express neurites, or can induce neurites to bifurcate (on average) only after a certain distance from the soma.

A wide variety of plausible arborization patterns emerge with this design (Senft & Ascoli 1999). However, the method used to identify the growth spots that are to receive additional segments is computationally inefficient, as it requires the sequential querying of large numbers of pre-existing segments. More importantly, several parameters used in this algorithm are not directly measurable from the experimental data. Thus, a potentially extensive parameter space must be searched to obtain the values of the statistical variables that best match the morphological details observed in specific neuronal classes.

In order to solve these problems, a second, more systematic and efficient algorithm was implemented in ARBORVITAE. The dealing process is still divided into growth epochs, but bifurcating and extending phases are not segregated in different epochs and the number of epochs can be determined at the group level. Each epoch is assigned with a certain quota of dendritic length. In this algorithm, growth can only occur at terminal branches, i.e. no previously added 'internal' segments can receive a second daughter. Within an epoch, each tip either terminates or receives two daughter branches (i.e. it bifurcates), depending on a termination probability specified at the group and epoch levels. If a tip terminates, it is removed from the set of 'active spots' for the rest of the epochs, and will remain a terminal point in the mature neuron. If a tip receives two daughter branches, two values for interbifurcation length are sampled, and subtracted from the epoch length quota. Each branch then extends for the sampled length with a number of shorter cylindrical subsegments, each with a length also specified at the group level (the shorter the length of the individual segments, the larger the number of segments to be recruited for appropriately elongating the branch). The end points of the two daughter branches are then added to the list of active spots. An epoch ends when the whole length quota is used. In this second algorithm, topological variability is provided by the random order in which active spots within a group are sampled for receiving daughter pairs. All the parameters used in this modified dealing process can be measured directly from experimental data under the assumption (reasonable for a

group) that growth epochs correspond roughly to physical distances around the soma (i.e. on average, dendrites grown earlier are closer to the soma). For example, the interbifurcation length, the total dendritic length within a certain distance from the soma, and the number of terminations and bifurcations versus path length, are all derivable from simple variations of Sholl analysis.

Many anatomical features of these 'intrinsic plans of growth', in both algorithms described above, can be modified in ARBORVITAE by 'environmental' variables, representing trophic or tropic influences. Environmental interactions are particularly consequential when neurons are generated in the context of larger networks: dendritic outgrowth can be modulated by the presence of afferent inputs, and axons can be made to navigate towards (or away from) a hierarchy of somatic and neuritic targets (Senft & Ascoli 1999). While such global interactions have their effects at the detailed level of individual segments (modulating their length and orientation), they are specified to the program parsimoniously as a matrix of interactions among source and target groups. Hence the specific information for creating axon pathways or synaptic sites is generated on-the-fly from stochastic variation of group-level constraints, and need not be specified *a priori* for each morphogenetic events. Consequently, the structures created by ARBORVITAE are dynamic, so that the developmental processes, such as cell migration, neurite growth or remodelling and signal propagation, can occur asynchronously and feed back on each other. However, it is important to underline that, once generated, these virtual structures are as detailed as if traced experimentally. Indeed, ARBORVITAE is able to intermingle traced and algorithmically generated data within the same networks, in a uniform internal format.

(c) *Other relevant algorithms and rules*

Although L-NEURON and ARBORVITAE implement a variety of local and global algorithms to generate dendritic morphology, there are other stochastic and statistical models that should be mentioned here. A global, mainly topological approach inspired Van Pelt and co-workers to study a series of increasingly complex models of dendritic morphology (Van Pelt & Uylings 1999). The key assumption in Van Pelt's models is that the bifurcating probability of a dendritic branch depends exponentially on the order of branching (i.e. the number of bifurcations that separate a branch from the soma). Bifurcations can occur at terminal branches or at 'internal' branches (i.e. branches that already terminate in a bifurcation). Thus, in the simplest model, dendritic 'growth' is completely characterized by the exponential relationship between bifurcation probability and branch order, and by the ratio of terminal versus internal growth. Additionally, more elaborate models consider the timing of bifurcation events, the length distribution of attaching branches, and the total number of available branches (Van Pelt & Uylings 1999). Van Pelt's models have been demonstrated to reproduce faithfully time-dependent growth in several morphological classes, also taking into account dendritic pruning and tree asymmetry (reviewed in Van Pelt & Uylings 1999). However, they lack a description (and a physical rationale) of branch diameters and diameter-related processes, such as tapering and diameter changes at bifurcations.

A different strategy has been proposed by Winslow *et al.* (1999) to generate dentate gyrus granule cell dendrites. In this study, the starting process is to reproduce the empirical distribution of the distances of all bifurcation points from the soma as normalized with respect to the farthest branch point. Then, branch points are connected starting from the soma by sampling the closest 'free' point from the growing 'tip', thus generating a one-dimensional projected tree. Next, this mono-dimensional tree is inflated into a 3D tree by stochastically projecting width and depth coordinates according to the empirical shape of the tree. Finally, coordinates are rescaled to reproduce the 'real' values, and diameter and dendritic wiggle are added, as sampled from empirical, distance-dependent distributions (Winslow *et al.* 1999). This model, which does not pretend to rely on any biophysical rationale, was not extensively characterized for its morphological realism, and was never tested for different morphological classes.

Additional local and global anatomical rules that can in principle be incorporated in morphological models of dendritic growth, such as the arbour optimization by Cherniak *et al.* (1999), were recently reviewed (Ascoli 1999). These additional or alternative constraints could be incorporated in L-NEURON and ARBORVITAE for direct empirical testing.

3. DATABASE STRUCTURE AND AVAILABILITY

We are using L-NEURON and ARBORVITAE to investigate systematically the potential of the 'computational neuroanatomy' approach for neuroscience databases. We have virtually generated anatomically plausible neurons for several morphological classes, including cerebellar Purkinje cells, hippocampal pyramidal cells and interneurons, dentate gyrus granule cells, thalamic relay neurons, cortical pyramidal and stellate cells, and spinal cord motor neurons (Senft 1997; Senft & Ascoli 1999; Ascoli 1999; Ascoli & Krichmar 2000). In this section, we describe a publicly available electronic database of virtual neurons that we are constructing as an example of the computational neuroanatomy potential (figure 2). This database, initially containing entries from the Purkinje cell and motor neuron morphological classes, is published through the Internet at http://www.krasnow.gmu.edu/L-Neuron/index.html (case sensitive).

The structure of the database is similar to that of an analogous archive of experimental dendritic morphologies (Cannon *et al.* 1998), which is also available electronically (and through links from the Internet address above). In particular, the database is formatted as a table. Each row of the table represents a single-neuron entry. The seven columns contain an entry name, neuron type (e.g. motor neuron or Purkinje cell), algorithm type (e.g. Hillman, ARBORVITAE, etc.), link to a comment file, link to a picture, link to an SWC file and links to other files (if any). The comment file is a simple text describing when and how the virtual neuron was generated, and providing references to its quantitative analysis, when available. The picture is an image of the generated neuron in portable format (usually JPEG). The SWC files are the digital

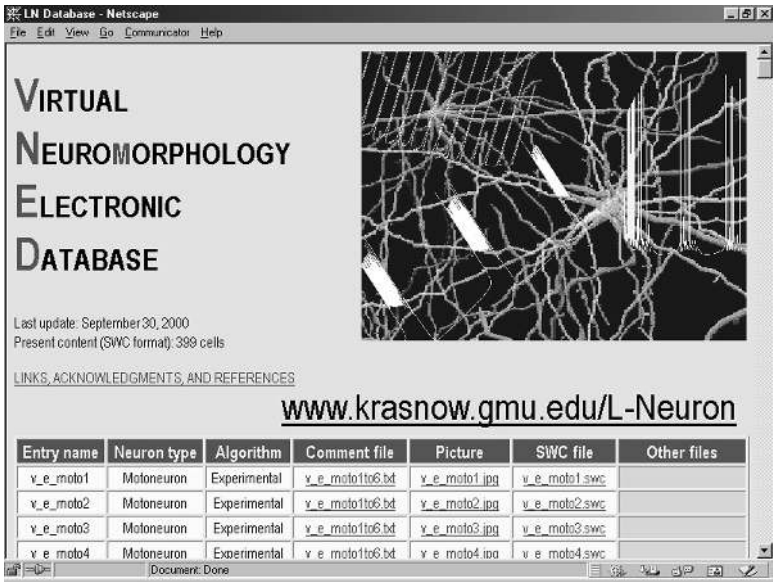


Figure 2. A screenshot of the virtual neuromorphology electronic database. The database is available online from the L-NEURON Web page, and contains pictures (JPEG) and SWC files of neuronal structures generated with a variety of local and global algorithms. The database also stores comment files, parameter files, links to conversion routines and references to the original experimental data.

anatomical data of the neurons, and constitute the core items of the database. The SWC format was also chosen in analogy with the above mentioned experimental database (Cannon *et al*. 1998). Neurons in SWC can be visualized three-dimensionally, rotated and zoomed through any Java-enabled internet browser (e.g. Netscape or Explorer in versions 4.0 or higher) with Cannon's applet Cvapp.

Additionally, for virtual neurons generated with L-NEURON, the L-NEURON source file and the binary VOL file are also available in the database (as hyperlinks in the 'other files' column). VOL neuron files, too, can be viewed three-dimensionally, rotated and zoomed with Lapré's software L-VIEWER, running under DOS and Windows. Both Cvapp and L-VIEWER are freely available (internet links provided in the database Web page). For virtual neurons generated with ARBORVITAE, the original output (in Eutectic-like format) is also available in the database.

The present structure of the database was designed to maximize simplicity and user friendliness. Only a very basic knowledge of internet browsing is required to scroll the page and select the links. However, more sophisticated options, such as online generation of neurons, advanced entry searches and comparisons, are not implemented at this time.

It is relevant to note here that the impact of this and other databases of neuronal morphology on the neuroscientific community will also depend on the availability of simple software tools that can extract flexibly a variety of user-defined morphological parameters from the archives. We have developed one such program called L-MEASURE (Scorcioni *et al*. 2000, 2001), which is fully compatible with SWC and most other formats (including Eutectic and Neurolucida). L-MEASURE is also publicly available (http://www.krasnow.gmu.edu/L-Neuron/index.html, case sensitive), but at present it does not allow the online analysis of archived data (both the software and the data must be downloaded onto the local disk).

The experimental dataset for motor neuron morphology is constituted by six gastrocnemius α-motor neurons from the spinal cord of adult cats. These neurons were labelled by intracellular injection of horseradish peroxidase, completely reconstructed from serial sections and digitized into computer files by Cullheim *et al*. (1987). The original computer file format is a redundant union of the SWC-like and Eutectic-like formats described in § 1b of this paper. More precisely, the dendritic trees are represented as a series of cylindrical compartments, each described by a line of thirteen values corresponding to a label, branch order, branch number, segment number, segment type (extending, bifurcating or terminating), x, y, z positions of the cylinder starting point, and x, y, z positions of the cylinder ending point, segment length and segment diameter. The numbering system for branches and segments is described in detail in Cullheim *et al*. (1987). The original digital files were converted to SWC formats with a simple routine written in AWK (Aho *et al*. 1988) and running under Unix. The six experimental motor neurons in SWC formats as well as the conversion routines are available in the online database.

The experimental dataset for Purkinje cell morphology is constituted by three neurons from the albino guinea pig cerebellar cortex. These cells were labelled with horseradish peroxidase and morphologically reconstructed by Rapp *et al*. (1994). The original computer files in Eutectic format were also converted into SWC with an AWK routine running under Unix. The three SWC Purkinje cell files as well as the conversion routine are available at the database Internet site. It is useful to notice that, despite the limited number of neurons experimentally available in the Purkinje class, the high morphological complexity of these cells makes each entry extremely data-rich. For example, the three dendritic trees corresponding to the three Purkinje cells contain a total of 1308 bifurcations, while the 70 dendritic trees

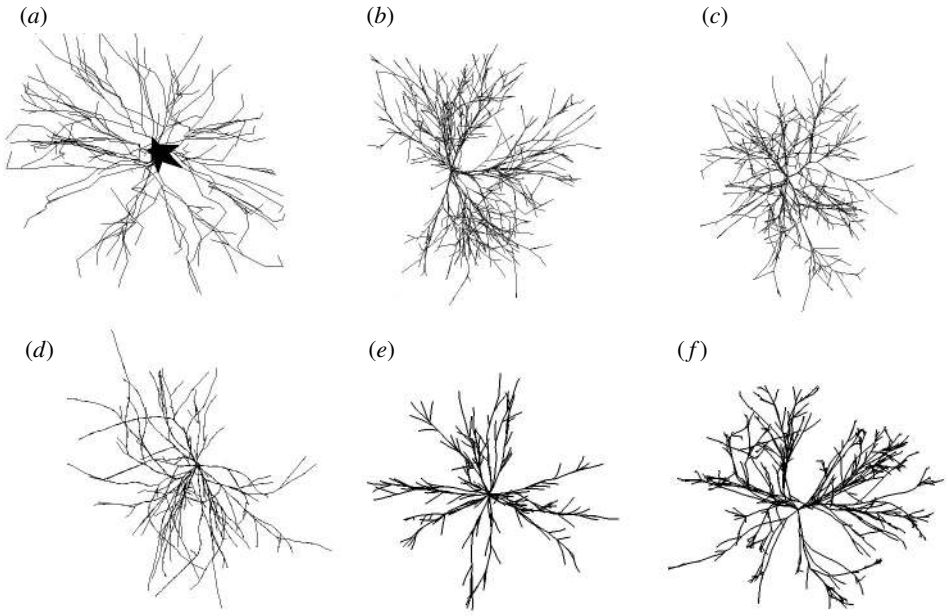


Figure 3. Examples of motor neurons from the database. (*a*) An experimental motor neuron (v_e_moto1) from Cullheim *et al.* (1987). (*b*) An L-NEURON cell generated with Hillman's algorithm, 'poliko' option (v_hp_moto1). (c) A cell generated as in (*b*) but using Tamori's variation (v_tp_moto1). (*d*) An L-NEURON cell generated with Burke's algorithm (v_b_moto1). Parameters were taken from Burke *et al.* (1992), or, when not available, measured from the experimental data. (*e*) An ARBORVITAE cell generated using 'algorithm one' (v_a1_moto1). (*f*) An ARBORVITAE cell generated using 'algorithm two' (v_a2_moto1).

corresponding to the six motor neurons contain a total of 950 bifurcations.

Virtual neurons were generated by L-NEURON and ARBORVITAE using a variety of algorithms and settings. For each algorithm, one to ten sets of six virtual motor neurons and three virtual Purkinje cells were created. Within each group, different random seeds were used in the generation of individual neurons. In particular, the following algorithms (described in sections 2a and 2b of this paper) were used:

- L-NEURON, Hillman-like algorithm;
- L-NEURON, Hillman-like algorithm with 'poliko' implementation;
- L-NEURON, Hillman-like algorithm, Tamori variant, with 'poliko' implementation;
- L-NEURON, Burke's algorithm;
- L-NEURON, miscellaneous algorithms with tropism (low value);
- L-NEURON, miscellaneous algorithms with tropism (high value);
- ARBORVITAE, first algorithm;
- ARBORVITAE, second algorithm;
- ARBORVITAE, second algorithm with trophic gradient (low value);
- ARBORVITAE, second algorithm with trophic gradient (high value).

The electronic database of virtual morphology currently contains 390 generated neurons (plus nine experimentally acquired neurons). More virtual neurons are going to be continuously added to this initial collection as new algorithms and different morphological classes are tested.

Figure 3 displays pictures of five virtual motor neurons generated with different algorithms as well as one experimental motor neuron. Even within the same morphological class (or a given set of parameter distributions and choice of algorithm) no two neurons are identical: the neurons shown in figure 3 are just typical examples from their respective groups.

Analogously, figure 4 displays pictures of five virtual (and one real) Purkinje cells generated with the same algorithms used in figure 3. Each of the algorithms can generate structures as different as motor neurons and Purkinje cells by using different statistical distributions of morphological parameters.

Figure 5 shows the effect of tropism on dendritic morphology. Here, real neurons are compared with virtual neurons generated with various algorithms with moderate or excessive tropism.

4. VERIFICATION AND VALIDATION OF COMPUTATIONAL NEUROANATOMY ALGORITHMS

An important aspect in computational neuroanatomy is the validation of algorithms. A comparison of the multiple entries of the database of virtual morphology (even as sketchily summarized in figures 2, 3 and 4) makes it immediately apparent that groups of neurons created with different algorithms are different. This structural variability is conceptually unlike both the natural variability observed within a morphological class (or within a group of virtual neurons generated stochastically with the same algorithm) and the morphological differences distinguishing two anatomical classes of neurons. How can one decide which algorithm is the best?

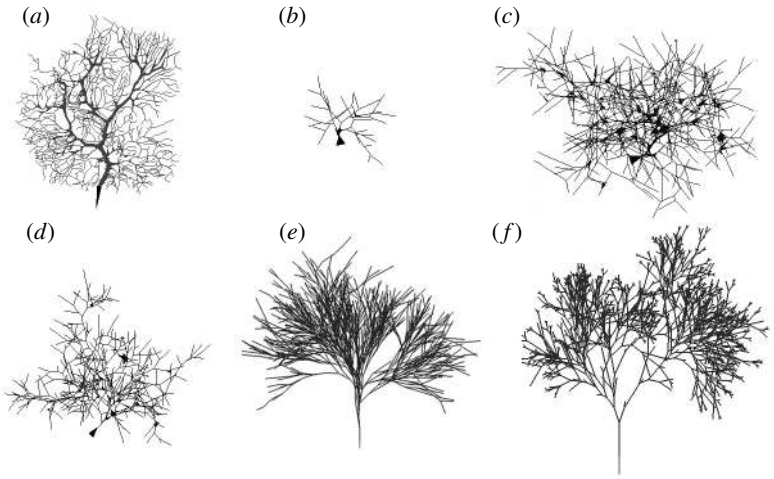


Figure 4. Examples of Purkinje cells from the database. (*a*) An experimental Purkinje cell (v_e_purk1) from Rapp *et al.* (1994). (*b*) An L-NEURON cell generated with Hillman's algorithm, without 'poliko' option (v_h_purk1). (*c*) Same as in (*b*), but with 'poliko' option, which remarkably increases the number of segments grown (v_hp_purk1). (*d*) Same as in (*c*), but using Tamori's variation, which improves the dendritic orientation (v_tp_purk1). (*e*) An ARBORVITAE cell generated using 'algorithm one' (v_a1_purk1). (*f*) An ARBORVITAE cell generated using 'algorithm two' (v_a2_purk1).

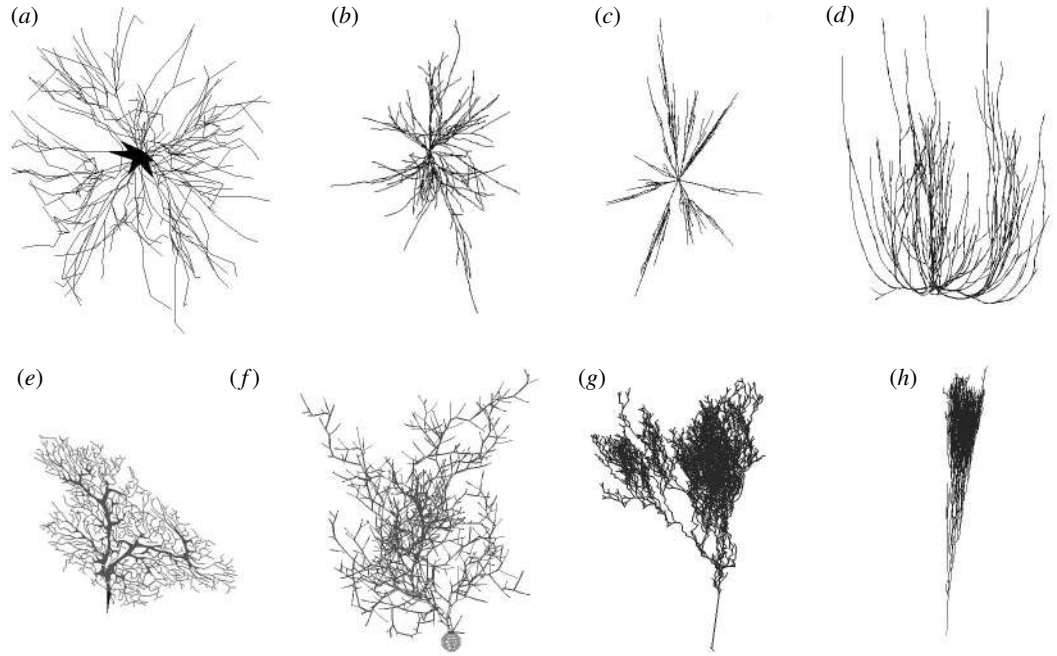


Figure 5. Effect of tropism on dendritic structure. (*a*) Another experimental motor neuron (v_e_moto2) from Cullheim *et al.* (1987). (*b*) An L-NEURON cell generated with Burke's algorithm, using a moderate amount of tropism, pushing dendrites away from the soma (v_bt_moto1). (*c*) Same as in (*b*), with an excessive amount of tropism (v_bts_moto1). (*d*) Same as in (*c*), setting tropism in the *y* axis (vertical on the page) instead of away from the soma (v_bty_moto1). (*e*) Another experimental Purkinje cell (v_e_purk2) from Rapp *et al.* (1994). (*f*) An L-NEURON cell generated using Hillman's algorithm with 'poliko option' and a moderate amount of tropism, both away from the soma and in the *y* axis (v_hpt_purk1). Notice the remarkable difference between this cell and the corresponding cell without tropism (figure 4*c*). (*g*) An ARBORVITAE cell generated using 'algorithm two' and tropism away from the soma (v_a2tl_purk1). (*h*) Same as in (*g*) with greater tropism (v_a2th_purk1): compare with the 'regular' cell, figure 4*f*.

There are several levels at which algorithms to generate dendritic morphology can be validated. First, one has to check that dendritic structures are being generated according to the parameters specified by the algorithm and by the user. A simple way to do this is to measure from the generated neurons the morphological parameters that the algorithm used in the generation (called 'basic parameters'). For example, using Hillman's algorithm one could measure the interbifurcation length of each branch in the group of virtual neurons. The statistical distribution of these values should correspond to the statistical distribution fed to the program to generate the dendrites. In order for two distributions of values to be statistically equivalent, they need to have the same

moments. Statistical distributions used in our algorithms for fitting the experimental data are fully characterized by the first two moments, i.e. mean and variance. Some distributions used in the algorithms, like uniform or gamma distribution, are conventionally specified with their standard parameters, however from these one can derive the first two moments (additionally, distributions in L-NEURON and ARBORVITAE are truncated at minimum and maximum boundaries). Thus, in this case, the problem of algorithm validation reduces to testing the 'null' hypothesis that the empirical distributions of the basic parameters collected from the set of virtual neurons are consistent with the distributions of these same parameters used for their generation. The statistical distributions used in the generation algorithms are therefore treated as known population probability distributions, and standard statistical tests are performed in order to assess the hypothesis. The results of testing are typically reflected in the p-value, stating the confidence level in accepting the null hypothesis.

However, it is necessary to note that statistical testing provides only an indirect algorithm validation. In fact, even if the test of the null hypothesis is positive, it only asserts that the probability of the data being sampled from a different population distribution is sufficiently low. Thus, although negative results of statistical testing would strongly indicate errors in the algorithm implementation, positive testing can never formally establish the correctness of the algorithm.

Assuming that the algorithm is in fact generating the correct parameters with the specified distributions, one needs to test the accuracy of the algorithmic description of a morphological class. In other words, the group of virtual neurons has to be compared with the corresponding group of real neurons. This can be done superficially through visual inspection. A trained neuroanatomist can quite accurately spot subtle differences between groups of cells. Thus, how anatomically plausible virtual neurons look is usually a good starting point to verify whether an algorithm is 'good'. A more quantitative way to compare groups of real and virtual neurons is to evaluate the similarity of the statistical distributions of morphological parameters other than those trivially used by the algorithm (called 'emergent parameters', as opposed to basic parameters). Statistical considerations similar to those discussed for the validation of the implementation apply here.

Emergent morphological parameters can be chosen at any level of complexity. Neurons can be characterized by single scalar numbers (e.g. total dendritic length), or by how local measures vary with location (e.g. dendritic diameter versus distance from the soma along the path) or with any other local measure (e.g. distribution of branch surface area versus daughter diameter ratio). In fact any correlation between morphological parameters (discussed in section 1b of this paper) can be used as a source of emergent parameters. Morphological parameters or distributions that are 'basic' for one algorithm can be considered 'emergent' for a different algorithm.

In general, as many emergent parameters as possible should be tested. If there is a statistically significant discrepancy in the comparison of any emergent parameter between real and virtual neurons, then the morphological algorithm used to generate those virtual neurons is not fully accurate. However, the discrepancy can provide useful anatomical feedback to modify and improve the algorithm. The complete morphological analysis of all the groups of virtual neurons collected in the online morphological database is beyond the scope of this discussion and will be published elsewhere (Ascoli *et al.* 2001).

5. CONCLUDING REMARKS

An important goal in computational neuroanatomy at the single-cell level consists of creating algorithms to generate virtual dendritic structures that are morphologically equivalent to real neuronal dendrites. To the best of our knowledge, L-NEURON and ARBORVITAE are the first two programs that implement in software neuroanatomical algorithms to generate complete virtual dendrites (other algorithms, while constituting influential scientific advances, have not been implemented to generate complete 3D structures; see, for example, Van Pelt & Uylings (1999).

In principle, an algorithm that can describe successfully the morphology of a mature neuron does not necessarily also reproduce the intermediate stages of development. In other words, the sequence of algorithmic steps does not always correspond to biological growth stages, even if the empirical description of the final product is excellent. Ideally, an algorithm that tries to reflect developmental mechanisms faithfully should also describe complex phenomena such as dendritic pruning and synaptic elimination. Indeed, it might be impossible to simulate dendritic growth processes accurately without embedding the neuron in the whole (simulated) network. If this proves to be the case, then only a modelling tool as complex as ARBORVITAE, where developmental interactions between anatomy and physiology can in principle be simulated, could accurately describe neuronal growth.

At this early stage of computational neuroanatomy, however, even unrealistic assumptions in the growth mechanism can constitute a useful starting point in morphological modelling. In fact, the analysis of the first 'generation' of products can help refine hypotheses and thus build an improved model to obtain a second generation of products. We are hopeful that this process would converge to an accurate algorithm describing and generating neuronal morphology. The attempt to relate systematically the discrepancies between virtual and real data to the logical steps of the algorithm is greatly facilitated if the algorithmic basic parameters correspond to intuitively accessible biophysical correlates. As previously mentioned in this paper, this is the case for some algorithms (e.g. Hillman's, see § 2a) but not others (e.g. Winslow's, see § 2c).

Both local and global algorithms implemented in L-NEURON and ARBORVITAE are extremely promising in the search for an accurate computational description of neuronal morphology. A large base of starting models is particularly useful in view of the possibility that a certain algorithm could describe perfectly a few morphological classes, but not others, and vice versa. From the user's point of view, local and global algorithms offer somewhat complementary advantages. Local algorithms are

conceptually more intuitive and their basic parameters are generally easier to measure from experimental data. Global algorithms are more flexible and potentially more detailed in the description of the growth processes. These differences are also reflected in the complexity of the implementations, whereas the L-NEURON program has a simpler structure than ARBORVITAE.

The potential advantages of the algorithmic description of neuronal structure are remarkable. If L-NEURON and ARBORVITAE, by measuring values of fundamental parameters from a restricted number of available experimental neurons, can generate virtual neurons that are anatomically indistinguishable from the real ones, two extremely important features for morphological databases, namely data compression and amplification, could be achieved. Data compression derives from the ability to describe quantitatively and completely any number of neurons from a morphological class with just a few statistical distributions of fundamental parameters. Data amplification is an intrinsic property of the computational neuroanatomy approach because, from a fixed number of experimental neurons, any (arbitrarily large) number of non-identical virtual analogues can be generated.

The electronic database described in this paper provides a source of clear examples to illustrate these aspects and possibilities. The six experimental motor neurons in uncommented SWC format occupy 63.7 kb of disk space, once compressed. Analogously, the size of the three experimental Purkinje cells (also in compressed, uncommented SWC format) is 47.1 kb. In contrast, the L-NEURON descriptor files for these groups range from 289 and 355 bytes (plain Hillman parameters occupying the least amount of space, and Burke parameters with tropism occupying the most).

A mammalian brain contains between 10^{11} and 10^{13} neurons and an estimated 500–2000 morphological classes (Williams & Herrup 1998; Shankle *et al*. 1998). In order to store a neuroanatomical database containing single-neuron morphological data for an entire such organ, an archive size of approximately 1000 Tb would be necessary (this estimate is completely theoretical, since it would be impossible to acquire such an amount of experimental data given the current technical limitations discussed at the beginning of this paper, § 1a). In contrast, starting from the experimental data acquired from a small fraction of the neurons in each morphological class, algorithms such as those implemented in L-NEURON and ARBORVITAE would enable the statistical description of dendritic morphology for an entire brain in a few megabytes of disk space.

The next crucial step in this line of research is certainly the extensive, quantitative, morphological characterization of virtual neurons, and their comparison with real cells of the corresponding morphological classes. Furthermore, additional morphological classes (particularly, hippocampal neurons) will be similarly modelled and analysed. This will help us investigate whether a single algorithm can be constructed that accurately describes several (or all) morphological classes.

This work was supported in part by Human Brain Project grant R01-NS39600, funded jointly by the National Institute of Neurological Disorders and Stroke and the National Institute of Mental Health (National Institutes of Health). We are thankful to Robert E. Burke and Moshe Rapp for making the experimental morphological data of motor neurons and Purkinje cells, respectively, available to us.

REFERENCES

Aho, A. V., Kernighan, B. W. & Weinberger, P. J. 1988 *The AWK programming language*. New York: Addison-Wesley.

Ascoli, G. A. 1999 Progress and perspective in computational neuroanatomy. *Anat. Rec.* **257**, 195–207.

Ascoli, G. A. & Krichmar, J. L. 2000 L-Neuron: a modeling tool for the efficient generation and parsimonious description of dendritic morphology. *Neurocomputing* **32–33**, 1003–1011.

Ascoli, G. A., Krichmar, J. L., Scorcioni, R., Nasuto, S. J. & Senft, S. L. 2001 Computer generation and quantitative morphometric analysis of virtual neurons. *Anat. Embryol*. (In the press.)

Burke, R. E., Marks, W. B. & Ulfhake, B. 1992 A parsimonious description of motor neurons dendritic morphology using computer simulation. *J. Neurosci*. **12**, 2403–2416.

Cabirol-Pol, M. J., Mizrahi, A., Simmers, J. & Meyrand, P. 2000 Combining laser scanning confocal microscopy and electron microscopy to determine sites of synaptic contact between two identified neurons. *J. Neurosci. Meth*. **97**, 175–181.

Cajal, S. Ramón y. 1894–1904 *Textura del Sistema Nervioso del Hombre y los Vertebrados* (1994 English translation N. Swanson & L. W. Swanson). Oxford University Press.

Cannon, R. C., Turner, D. A., Pyapali, G. K. & Wheal, H. V. 1998 An on-line archive of reconstructed hippocampal neurons. *J. Neurosci. Meth*. **84**, 49–54.

Cannon, R. C., Wheal, H. V. & Turner, D. A. 1999 Dendrites of classes of hippocampal neurons differ in structural complexity and branching patterns. *J. Comp. Neurol*. **413**, 619–633.

Capowski, J. J. & Schneider, S. A. 1985 A simple motor controller for computer-assisted microscopy. *J. Neurosci. Meth*. **13**, 97–102.

Cherniak, C., Changizi, M. & Kang, D. W. 1999 Large-scale optimization of neuron arbors. *Phys. Rev.* E **59**, 6001–6009.

Claiborne, B. J., Amaral, D. G. & Cowan, W. M. 1990 Quantitative, three-dimensional analysis of granule cell dendrites in the rat dentate gyrus. *J. Comp Neurol*. **302**, 206–219.

Cohen, M. J., Pitman, R. M. & Tweedle, C. D. 1973 Cobalt staining used for demonstrating neuronal branching. *J. Physiol*. **232**, 58P–59P.

Cohen, A. R., Roysam, B. & Turner, J. N. 1994 Automated tracing and volume measurements of neurons from 3-D confocal fluorescence microscopy data. *J. Microsc*. **173**, 103–114.

Cullheim, S., Fleshman, J. W., Glenn, L. L. & Burke, R. E. 1987 Membrane area and dendritic structure in type-identified Triceps Surae alpha motor neurons. *J. Comp. Neurol*. **255**, 68–81.

Dailey, M. E. & Smith, S. J. 1994 Spontaneous Ca^{2+} transients in developing hippocampal pyramidal cells. *J. Neurobiol*. **25**, 243–251.

De Schutter, E. 1999 Using realistic models to study synaptic integration in cerebellar Purkinje cells. *Rev. Neurosci*. **10**, 233–245.

Eilers, J. & Konnerth, A. 1997 Dendritic signal integration. *Curr. Opin. Neurobiol*. **7**, 385–390.

Glaser, J. R. & Glaser, E. M. 1990 Neuron imaging with Neurolucida—a PC-based system for image combining microscopy. *Comput. Med. Imag. Graph*. **14**, 307–317.

Godement, P., Vanselow, J., Thanos, S. & Bonhoeffer, F. 1987 A study in developing visual systems with a new method of staining neurons and their processes in fixed tissue. *Development* **101**, 697–713.

Gong, W. & Bertrand, G. 1990 A simple parallel 3D thinning algorithm. In *Proc. Inst. Electrical Electronic Engs 10th Int. Conf. Pattern Recognit.*, pp. 188–190.

Gustafsson, M. G., Agard, D. A. & Sedat, J. W. 1999 I5M: 3D widefield light microscopy with better than 100 nm axial resolution. *J. Microsc.* **195**, 10–16.

Haglund, M. M., Ojemann, G. A. & Hochman, D. W. 1992 Optical imaging of epileptiform and functional activity in human cerebral cortex. *Nature* **358**, 668–671.

Hillman, D. E. 1979 Neuronal shape parameters and substructures as a basis of neuronal form. In *The neurosciences, fourth study program* (ed. F. Schmitt), pp. 477–498. Cambridge, MA: MIT Press.

Hillman, D. E. 1988 Parameters of dendritic shape and substructure: intrinsic and extrinsic determination? In *Intrinsic determinants of neuronal form and function* (ed. R. J. Lasek & M. M. Black), pp. 83–113. New York: Liss.

Inoue, S. 1988 Progress in video microscopy. *Cell Motil. Cytoskeleton* **10**, 13–17.

Ishizuka, N., Cowan, W. M. & Amaral, D. G. 1995 A quantitative analysis of the dendritic organization of pyramidal cells in the rat hippocampus. *J. Comp. Neurol.* **362**, 17–45.

Krichmar, J. L., Nasuto, S. J., Scorcioni, R., Washington, S. D. & Ascoli, G. A. 2001 Effect of dendritic morphology on CA3 pyramidal cell electrophysiology. (Submitted.)

Larkman, A. U. 1991 Dendritic morphology of pyramidal neurons of the visual cortex of the rat: II. parameter correlations. *J. Comp. Neurol.* **306**, 320–331.

Luskin, M. B., Pearlman, A. L. & Sanes, J. R. 1988 Cell lineage in the cerebral cortex of the mouse studied in vivo and in vitro with a recombinant retrovirus. *Neuron* **1**, 635–647.

McNally, J. G., Karpova, T., Cooper, J. & Conchello, J. A. 1999 Three-dimensional imaging by deconvolution microscopy. *Methods* **19**, 373–385.

Muller, K. J. & McMahan, U. J. 1976 The shapes of sensory and motor neurons and the distribution of their synapses in ganglia of the leech: a study using intracellular injection of horseradish peroxidase. *Proc. R. Soc. Lond.* B **194**, 481–499.

Muller, T. 1989 Local supravital staining with methylene blue as marking technique for neurons. *Z. Mikrosk. Anat. Forsch.* **103**, 967–970.

Nowakowski, R. S., Hayes, N. L. & Egger, M. D. 1992 Competitive interactions during dendritic growth: a simple stochastic growth algorithm. *Brain Res.* **576**, 152–156.

Oldenbourg, R., Terada, H., Tiberio, R. & Inoue, S. 1993 Image sharpness and contrast transfer in coherent confocal microscopy. *J. Microsc.* **172**, 31–39.

Oldenbourg, R. & Mei, G. 1995 New polarized light microscope with precision universal compensator. *J. Microsc.* **180**, 140–147.

Orbach, H. S., Cohen, L. B. & Grinvald, A. 1985 Optical mapping of electrical activity in rat somatosensory and visual cortex. *J. Neurosci.* **5**, 1886–1895.

Pawley, J. B. (ed.) 1990 *Handbook of biological confocal microscopy.* New York: Plenum Press.

Piston, D. W. 1999 Imaging living cells and tissues by two-photon excitation microscopy. *Trends Cell. Biol.* **9**, 66–69.

Price, J., Turner, D. & Cepko, C. 1987 Lineage analysis in the vertebrate nervous system by retrovirus-mediated gene transfer. *Proc. Natl Acad. Sci. USA* **84**, 156–160.

Prusinkiewicz, P. & Lindenmayer, A. 1990 *The algorithmic beauty of plants.* New York: Springer Verlag.

Purves, D. & Hadley, R. D. 1985 Changes in the dendritic branching of adult mammalian neurons revealed by repeated imaging in situ. *Nature* **315**, 404–406.

Rapp, M., Segev, I. & Yarom, Y. 1994 Physiology, morphology, and detailed passive models of guinea-pig cerebellar Purkinje cells. *J. Physiol.* **474**, 101–118.

Rizzuto, R., Carrington, W. & Tuft, R. A. 1998 Digital imaging microscopy of living cells. *Trends Cell. Biol.* **8**, 288–292.

Roy, N. S., Benraiss, A., Wang, S., Fraser, R. A., Goodman, R., Couldwell, W. T., Nedergaard, M., Kawaguchi, A., Okano, H. & Goldman, S. A. 2000 Promoter-targeted selection and isolation of neural progenitor cells from the adult human ventricular zone. *J. Neurosci. Res.* **59**, 321–331.

Samuels, D. C., Hentschel, H. G. & Fine, A. 1996 The origin of neuronal polarization: a model of axon formation. *Phil. Trans. R. Soc. Lond.* B **351**, 1147–1156.

Schierenberg, E., Cole, T., Carlson, C. & Sidio, W. 1986 Computer-aided three-dimensional reconstruction of nematode embryos from EM serial sections. *Exp. Cell. Res.* **166**, 247–252.

Scorcioni, R., Krichmar, J. L. & Ascoli, G. A. 2000 Algorithmic extraction of morphological statistics from electronic archives of neuroanatomy. *J. Neurosci.* **S26** (Suppl.), 188.

Scorcioni, R. & Ascoli, G. A. 2001 Algorithmic extraction of morphological statistics from electronic archives of neuroanatomy. *Lect. Notes Comput. Sci.* (In the press.)

Senft, S. L. 1995 Derivation of neuron geometry from confocal scans. *J. Neurosci.* **S21** (Suppl.), 1078.

Senft, S. L. 1997 A statistical framework to present developmental neuroanatomy. In *Biobehavioral foundations* (ed. J. Donahoe). Elsevier.

Senft, S. L. & Ascoli, G. A. 1999 Reconstruction of brain networks by algorithmic amplification of morphometry data. *Lect. Notes Comput. Sci.* **1606**, 25–33.

Shankle, W. R., Romney, A. K., Landing, B. H. & Hara, J. 1998 Developmental patterns in the cytoarchitecture of the human cerebral cortex from birth to 6 years examined by correspondence analysis. *Proc. Natl Acad. Sci. USA* **95**, 4023–4028.

Shepherd, G. M. & Harris, K. M. 1998 Three-dimensional structure and composition of CA3→CA1 axons in rat hippocampal slices: implications for presynaptic connectivity and compartmentalization. *J. Neurosci.* **18**, 8300–8310.

Shoham, D., Glaser, D. E., Arieli, A., Kenet, T., Wijnbergen, C., Toledo, Y., Hildesheim, R. & Grinvald, A. 1999 Imaging cortical dynamics at high spatial and temporal resolution with novel blue voltage-sensitive dyes. *Neuron* **24**, 791–802.

Sholl, D. A. 1953 Dendritic organization of the neurons of the visual and motor cortices of the cat. *J. Anat.* **87**, 387–406.

Sims, S. J. & Macagno, E. R. 1985 Computer reconstruction of all the neurons in the optic ganglion of *Daphnia magna*. *J. Comp. Neurol.* **233**, 12–29.

Somogyi, P. 1990 Synaptic connections of neurons identified by Golgi impregnation: characterization by immunocytochemical, enzyme histochemical, and degeneration methods. *J. Electron Microsc. Tech.* **15**, 332–351.

Soto, G. E., Young, S. J., Martone, M. E., Deerinck, T. J., Lamont, S., Carragher, B. O., Hama, K. & Ellisman, M. H. 1994 Serial section electron tomography: a method for three-dimensional reconstruction of large structures. *NeuroImage* **1**, 230–243.

Stearns, T. 1995 Green fluorescent protein. The green revolution. *Curr. Biol.* **5**, 262–264.

Stevens, J. K., Davis, T. L., Friedman, N. & Sterling, P. 1980 A systematic approach to reconstructing microcircuitry by electron microscopy of serial sections. *Brain Res.* **2**, 265–293.

Stewart, W. W. 1981 Lucifer dyes—highly fluorescent dyes for biological tracing. *Nature* **292**, 17–21.

Tamori, Y. 1993 Theory of dendritic morphology. *Phys. Rev.* E **48**, 3124–3129.

Tsodyks, M., Kenet, T., Grinvald, A. & Arieli, A. 1999 Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* **286**, 1943–1946.

Uylings, H. B. M., Ruiz-Marcos, A. & Van Pelt, J. 1986 The metric analysis of three-dimensional dendritic tree patterns: a methodological review. *J. Neurosci. Meth.* **18**, 127–151.

Van Ooyen, A., Willshaw, D. J. & Ramakers, G. J. A. 2000 Influence of dendritic morphology on axonal competition. *Neurocomputing.* (In the press.)

Van Pelt, J. & Uylings, H. B. M. 1999 Natural variability in the geometry of dendritic branching patterns. In *Modeling in the neurosciences—from ionic channels to neural networks*. (ed. R. R. Poznanski), **4**, 79–108. Amsterdam, The Netherlands: Harwood Academic Publishers.

Van Pelt, J., Uylings, H. B. M., Verwer, R. W. H., Pentney, R. J. & Woldenberg, M. J. 1992 Tree asymmetry—A sensitive and practical measure for binary topological trees. *Bull. Math. Biol.* **54**, 759–784.

White, E. L., Amitai, Y. & Gutnick, M. J. 1993 A novel approach to correlative studies of neuronal structure and function. *Israel J. Med. Sci.* **29**, 549–555.

Williams, R. W. & Herrup, K. 1988 The control of neuron number. *A. Rev. Neurosci.* **11**, 423–453.

Winslow, J. L., Jou, S. F., Wang, S. & Wojtowicz, J. M. 1999 Signals in stochastically generated neurons. *J. Comput. Neurosci.* **6**, 5–26.

Wong, W. T. & Wong, R. O. 2000 Rapid dendritic movements during synapse formation and rearrangement. *Curr. Opin. Neurobiol.* **10**, 118–124.