Modeling: From Ions to Networks (Koch, C. and Segev, I., eds), pp. 251–291, MIT Press
43 Jack, J.J.B., Noble, D. and Tsien, R.W. (1975) *Electric Current Flow in Excitable Cells*, Clarendon Press
44 Koch, C. *Biophysics of Computation: Information Processing in Single Neurons*, Oxford University Press (in press)
45 Koch, C. and Segev, I., eds (1998) *Methods in Neuronal Modeling: From Ions to Networks*, MIT Press
46 McKenna, T., Davis, J. and Zornetzer, S.F., eds (1992) *Single Neuron Computation*, Academic Press
47 Rall, W. *et al.* (1992) *Physiol. Rev.* 72, S159–S186
48 Rall, W. and Shepherd, G.M. (1968) *J. Neurophysiol.* 31, 884–915
49 Rall, W. (1967) *J. Neurophysiol.* 30, 1138–1168
50 Jack, J.J.B. *et al.* (1971) *J. Physiol.* 215, 353–380
51 Mendell, L.M. and Henneman, E. (1971) *J. Neurophysiol.* 34, 171–187
52 Rall, W. *et al.* (1967) *J. Neurophysiol.* 30, 1169–1193
53 Redman, S. and Walmsley, B. (1983) *J. Physiol.* 343, 117–133
54 Agmon-Snir, H. and Segev, I. (1993) *J. Neurophysiol.* 70, 2066–2085
55 Rall, W. and Rinzel, J. (1973) *Biophys. J.* 13, 648–688
56 Rinzel, J. and Rall, W. (1974) *Biophys. J.* 14, 759–790
57 Zador, A.M., Agmon-Snir, H. and Segev, I. (1995) *J. Neurosci.* 15, 1669–1682
58 Rall, W. *et al.* (1966) *Exp. Neurol.* 14, 44–56
59 Dodge, F.A. and Cooley, J.W. (1973) *IBM J. Res. Devel.* 17, 219–229
60 Bunow, B., Fleshman, J. and Segev, I. (1985) *Biol. Cybern.* 53, 41–56
61 Moore, J.W. *et al.* (1983) *J. Physiol.* 336, 301–311
62 Goldstein, S.S. and Rall, W. (1974) *Biophys. J.* 14, 731–757
63 Khodorov, B.I. and Timin, Ye N. (1975) *Prog. Biophys. Mol. Biol.* 30, 145–184
64 Manor, Y., Koch, C. and Segev, I. (1991) *Biophys. J.* 60, 1424–1437
65 Parnas, I. and Segev, I. (1979) *J. Physiol.* 295, 323–343
66 Traub, R. *et al.* (1991) *J. Neurophysiol.* 66, 635–649
67 Pinsky, P. and Rinzel, J. (1994) *J. Comput. Neurosci.* 1, 39–60
68 Mainen, Z.F. and Sejnowski, T.J. (1996) *Nature* 382, 363–366
69 Borst, A. and Egelhaaf, M. (1994) *Trends Neurosci.* 17, 257–263
70 Mel, W.B. (1993) *J. Neurophysiol.* 70, 1086–1101
71 Segev, I. (1995) in *The Handbook of Brain Theory and Neuronal Networks* (Arbib, M.A., ed.), pp. 282–289, MIT Press
72 Softky, W. (1994) *Neuroscience* 58, 13–41
73 Wilson, C.J. (1995) *J. Comput. Neurosci.* 2, 91–116
74 Miller, J.P., Rinzel, J. and Rall, W. (1985) *Brain Res.* 325, 325–330
75 Perkel, D.H. and Perkel, D.J. (1985) *Brain Res.* 325, 331–335
76 Baer, S.M. and Rinzel, J. (1991) *J. Neurophysiol.* 65, 874–890
77 Rall, W. and Segev, I. (1987) in *Synaptic Function* (Edelman, G.M., Gall, E.E. and Cowan, W.M., eds), pp. 605–636, Wiley and Sons
78 Shepherd, G.M. *et al.* (1985) *Proc. Natl. Acad. Sci. U. S. A.* 82, 2192–2195
79 De Schutter, E. and Bower, J. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 4736–4740
80 Segev, I. and Rall, W. (1991) *Soc. Neurosci. Abstr.* 17, 1515
81 Rapp, M. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 11985–11990
82 Segev, I. and Rall, W. (1988) *J. Neurophysiol.* 60, 499–523
83 Colbert, C.M. and Johnston, D. (1996) *J. Neurosci.* 16, 6676–6686
84 Amitai, Y. *et al.* (1993) *Cereb. Cortex* 3, 26–38
85 Schwindt, P.C. and Crill, W.E. (1998) *J. Neurophysiol.* 79, 2432–2446
86 Turner, R.W. *et al.* (1991) *J. Neurosci.* 11, 2270–2280
87 Segal, M. (1995) *J. Neurophysiol.* 74, 484–488
88 Mainen, Z.F. *et al.* (1995) *Neuron* 15, 1427–1439
89 Murphy, T.H. *et al.* (1994) *Nature* 263, 529–532
90 Buzsaki, G. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 9921–9925
91 Kim, H.G., Beierlein, M. and Connors, B.W. (1995) *J. Neurophysiol.* 74, 1810–1815
92 Bernander, O. *et al.* (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 11569–11573
93 Shepherd, G.M. and Brayton, R.K. (1987) *Neuroscience* 21, 151–165

Acknowledgements
This work was supported by grants from the ONR and the Israeli Academy of Science. We thank M. Rapp, S. Redman and A. Thomson for their critical reviews of the article.

The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data

Gordon M. Shepherd, Jason S. Mirsky, Matthew D. Healy, Michael S. Singer, Emmanouil Skoufos, Michael S. Hines, Prakash M. Nadkarni and Perry L. Miller

Gordon M. Shepherd, Jason S. Mirsky and Michael S. Singer are in the Section of Neurobiology, and Matthew D. Healy, Emmanouil Skoufos, Prakash M. Nadkarni and Perry L. Miller are at the Center for Medical Informatics and Dept of Anesthesiology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA. Michael S. Hines is at the Dept of Computer Science, Yale University, New Haven, CT 06520, USA.

What is neuroinformatics? What is the Human Brain Project? Why should you care? Supported by a consortium of US funding agencies, the Human Brain Project aims to bring to the analysis of brain function the same advantages of Internet-accessible databases and database tools that have been crucial to the development of molecular biology and the Human Genome Project. The much greater complexity of neural data, however, makes this a far more challenging task. As a pilot project in this new initiative, we review some of the progress that has been made and indicate some of the problems, challenges and opportunities that lie ahead.

Trends Neurosci. (1998) 21, 460–468

THE HUMAN BRAIN PROJECT originated during the 1980s in discussions between neuroscientists and forward-looking program directors at the National Institutes of Health and the National Science Foundation[1]. They realized that the development of new technologies for creating databases and database search tools, and of electronic means for information exchange, was proceeding at a pace that outstripped the abilities of most neuroscientists to use these technologies. From the viewpoint of the funding agencies this was cause for concern, because these were the kinds of 'enabling technologies' that would allow neuroscientists to make much more efficient use of their data (and the agencies to get 'more bang for the buck').

PII: S0166-2236(98)01300-9

Box 1. Organizations sponsoring the Human Brain Project

National Institute of Mental Health
National Institute on Drug Abuse
National Science Foundation
National Institute on Aging
National Institute on Child Health and Human Development
National Institute on Deafness and Other Communication Disorders
National Library of Medicine
Office of Naval Research
National Aeronautics and Space Administration
Fogarty International Center
Department of Energy
National Institute on Alcohol Abuse and Alcoholism
National Heart, Lung, and Blood Institute
National Institute of Dental Research
National Institute of Neurological Disorders and Stroke
National Cancer Institute

In the traditional mode of experimental science, each laboratory generates its data to solve a problem, publishes a fraction of it, and moves on. The published data remain in the hard copy volume or reprint, inaccessible by electronic means; the unpublished data are lost forever. In the electronic age, this is outdated; in an age of limited national resources, it is an inefficient use of the taxpayers' dollars.

This classical mold was broken during the 1980s at the gene and protein level by the gene-sequencing community. Because of the simple linear nature of the data, genetic and protein sequences could easily be submitted, in parallel with journal publication, to central databases where they could be archived and made available for further analysis. This analysis is carried out by 'informatics tools', which are the programs that enable a database to be queried to generate new information, from simple questions such as identifying similarities, which might represent homologies, with any arbitrary sequence, to computationally intensive searches such as correlated mutation analysis[2]. These tools have become an integral part of carrying out research on genes and proteins, as essential to molecular biology as are its experimental methods. The success of the Human Genome Project in sequencing the genome for the human and for other species has therefore depended on both efficient experimental methods and efficient informatics tools. The necessary conditions that make possible the efficient electronic use of the sequence data include: simple linear data; quality control by refereed publication; requirement for submitting published sequences in central databases; automated data submission; automated data search tools.

In order to consider how to develop informatics tools that would be as effective for neuroscience data as those for gene and protein data, a committee, representing a broad range of neuroscientists and informatics specialists, was convened by the Institute of Medicine around 1990. After nearly two years of deliberation, it recommended[3] that a pilot program be set up to fund laboratories and laboratory groups that would combine experimental approaches with the development of electronic 'enabling technologies'. This envisaged a new research specialty of 'neuroinformatics'. Out of this came a funding program in 1993 that was unusual in several respects. First, a project would require not only experimental neuroscientists but also informatics scientists as integral members of the research team. Second, its funds would be derived from a wide range of federal agencies that support neuroscience research and wished to extend that support with the new methods (see Box 1).

Finally, it acquired the name Human Brain Project to indicate a vision complementary to the Human Genome Project, of providing informatics support to enable neuroscientists to carry out a complete mapping of the molecules, cells, circuits and systems in the human brain, in health and disease. Fortuitously, the emergence of the World Wide Web at about the same time provided the ideal means for linking the pilot projects and making their databases and tools accessible to each other and to neuroscientists everywhere.

It should be emphasized that the Human Brain Project, like the Human Genome Project, embraces all species. This recognizes the fundamental principle underlying modern biomedical research, that basic research on all species is needed to contribute to an understanding of the human and to improvements in human health.

The first five years

From the start it was recognized that neuroscience data are much more complex than sequence data; as noted by Peter Pearson, 'genome project informatics is trivial by comparison'[4]. The informatics problems in neuroscience are therefore far more daunting than those currently being addressed in molecular genetics or most other fields. Some indication of this complexity is indicated in Table 1, which shows that neuroscience data come at a number of organizational levels.

In addition, at each level, the data come from multiple disciplines, in two and three spatial dimensions, and can vary in a fourth dimension, time. Thus, whereas in the Human Genome Project mapping the gene is the primary goal, with the informatics aspects being of secondary importance, in the Human Brain Project the main problem is not only organizing the neuroscience data but also developing new informatics tools that can deal with the much more complex forms of data and the relationships between them.

The goal of the first five-year period of funding of pilot studies was therefore modest: to explore the feasibility of different approaches to these difficult problems. Some of the initial projects will be summarized briefly; Web addresses are provided in Box 2.

Human brain imaging

One large effort has been directed to brain images. Many large centers are actively pursuing positron emission tomography (PET), magnetic resonance imaging (MRI), functional MRI (fMRI) and magnetoencephalography (MEG), in addition to traditional electroencephalographic (EEG) mapping of brain activity. These studies generate large numbers of very large datasets (many gigabytes and terabytes of data) of image data. As expressed by Vincent Cerf, one of

TABLE 1. Levels of function

Levels of brain function	Types of experimental neuroscience data
Behavior	Performance quantitation, video monitoring, drug testing
Distributed systems	2D and 3D axon-tracing between regions, electrophysiological recordings (spike timing), brain imaging and 3D brain maps
Specific regions	2D and 3D cytoarchitectonics of layers and functional columns, transmitter-receptor localization, anatomical, physiological and metabolic maps
Nerve cells	3D cell morphology, 3D functional imaging, electrophysiological recordings of action-potential firing patterns and membrane currents
Neuronal compartments	3D imaging of axon terminals, growth cones, dendrites, dendritic spines, 3D localization of organelles and synaptic microcircuits
Microcircuits	3D fine structure and imaging of synaptic patterns, synaptic pharmacology, action-potential firing patterns and synaptic currents and potentials
Organelles	2D and 3D fine structure and molecular composition of synapses, mitochondria, microtubules, etc.; recordings of synaptic currents and potentials
Molecules	3D molecular models of receptors, channels, enzymes and structural proteins, molecular physiology and pharmacology of transmitters, modulators, hormones, guidance molecules, growth factors and gene-transcription factors
Genes	DNA and protein sequences

the pioneers in the development of the Internet: 'We are in a very infantile state…regarding the indexing, cataloging and searching of nontextual information… . It's one thing to find text words in documents that are alike; it's something else to find similar images that relate to each other'[5].

The first five years of the Human Brain Project have seen a concerted effort by a number of laboratories towards solving the informatics problems involved in comparing human brain images. These problems include: how to construct reference brain images; how to relate brain images from different patients and from different laboratories, a process known as 'warping'; how to exchange these large datasets efficiently between laboratories; how to make such archives of brain images available for further 'data mining'. An example of progress towards these goals is provided by the International Consortium for Brain Imaging (ICBM) (http://www.loni.ucla.edu/icbm), consisting of laboratories at UCLA, University of Texas San Antonio, Montreal Neurological Institute, Stanford University, Albert Einstein College of Medicine in New York and Heinrich Heine University in Germany. A typical study[6] has involved comparing a representative normal brain (Fig. 1A) with the brain of a patient with Alzheimer's disease (Fig. 1B). The authors describe an algorithm that calculates a high-dimensional volumetric warp that brings both 3D images into register with each other (Fig. 1C). This warp involves $384 \times 384 \times 256 \times 3$ (~0.1 billion) degrees of freedom. In the example shown, the cerebellum of the Alzheimer's patient is contracted, and more subtle variations in the posterior frontal and cingulate regions are seen (Fig. 1D).

Studies such as this one are addressing how to characterize quantitatively the relatedness of 3D images obtained in different subjects by different laboratories. These and other advances can make brain imaging much more accurate and make much better use of the data, so that there is reduced duplication of effort, increased sharing of data, and more effective collaborations between laboratories. The better use of resources is a significant gain when funding multiple imaging centers.

Brain maps and atlases

These types of studies logically require the development of brain atlases of normal brains for comparison with brains in different disease states. There have been several initiatives taken, inside and outside the Human Brain Project, in this direction. One of the earliest of these was Brain Browser[7] for the rat brain. Digital atlases are being developed for several species, including human[8,9]. These have utility for research, and also have become integral parts of the teaching of neuroanatomy to medical students and graduate students. Although the main aim of the Human Brain Project is to provide tools for neuroscience research, the added benefits for science education, including school-age children and the general public, will be of considerable value.

In addition to the need for brain atlases, the analysis of brain function requires the construction of maps of organization within regions. Leading the way are maps of the visual areas, which have become extremely complex, with dozens of sub-areas and their interconnections[10]. Mapping these areas requires converting the 3D surfaces of the visual cortical areas into 2D representations. Sophisticated software for doing the necessary conversion and warping has been developed (CARET) and is being made available on the Web (v1.wustl.edu/caret) so that it can be applied not only to visual cortex mapping but also to other areas.

Neuronal properties

An important commitment from the start of the Human Brain Project was to support informatics tools for data at all levels of brain organization. Thus, as shown in Table 1, between sequence data at the most basic level and brain images at the highest level are a number of levels in which neuronal membrane properties are crucial for the neural basis of brain function. Several pilot projects are dealing with data at these different levels, as indicated in Box 2. Thus, databases for the fine structure of synapses, neuronal membrane properties, neuronal morphology and physiological data are being constructed. Links are being built to other parallel efforts, such as the G Protein Coupled Receptor consortium (GPCRDB: www.swift.embl-heidelberg.de/7TM/), a database of 3D reconstructions of vestibular hair cells and their nerve terminals (biocomp.arc.nasa.gov/reconstructions) and the Ion Channel Network (a comprehensive database of membrane properties; www.le.ac.ak.csn), to name just a few examples. Links are also being constructed to public sequence databases such as GenBank and SwissProt, as well as to focused sequence databases such as for *Caenorhabditis elegans* (eatworms.swmed.edu).

Integrating data into models

In order to understand the significance of a particular property in a particular neuron, it is necessary to assess that property in relation to other properties at that site. Integrating data in this way leads naturally to the construction of computational models, which

Box 2. Pilot projects in neuroinformatics funded by the Human Brain Project

Neuronal properties

Neuronal vulnerability and informatics in human disease
F. Bloom, Scripps Research Institute
www-hbp-np.scripps.edu/home; www-hbp.scripps.edu
neurozoom.scripps.edu/ServerData/HTMLs/Software/NZ/Server/default.html
www-cajal.ucsd.edu; www.med.uci.edu/~hbp/index.html

Development of a three-dimensional cell-centered neural database
M. Ellisman, University of California, San Diego

Somatosensory cortical neuron physiology: a Web database
D. Gardner, Cornell University Medical College
cortex.med.cornell.edu

Three-dimensional structure and function of synapses in the brain
K. Harris, Children's Hospital, Harvard Medical School

Three-dimensional reconstruction of synapses
P. Luther, University of Maryland

Integration of multidisciplinary sensory data
G. Shepherd, Yale Medical School
senselab.med.yale.edu/ordb; senselab.med.yale.edu/neurondb;
senselab.med.yale.edu/modeldb

Neuronal and network models

Neural plasticity: data and computational structures
M. Arbib, University of Southern California
www-hbp.usc.edu/HBP/

A simulator-based neuronal database
J. Bower, California Institute of Technology
www.bbb.caltech.edu/hbp/

Parallel simulation of large-scale neuronal models
N. Goddard, Carnegie Mellon University

Brain maps and atlases

High-resolution brain atlas – data acquisition and user access
R. Sidman, Harvard University

Reconstructions and representations of cerebral cortex
D. van Essen, Washington University
v1.wustl.edu

Brain imaging

Extended thin-plate splines for brain variation in three dimensions
F. Bookstein, University of Michigan
ftp://brainmap.med.umich.edu/pub/edgewarp

Structural information framework for brain mapping
J. Brinkley, University of Washington
www1.biostr.washington.edu/brain project

Advanced methods for neuroimaging data analysis
J. Cohen, University of Pittsburgh

Mapping brain function by combined MRI, MEG and fMRI
J. George, Los Alamos National Laboratory

Goal-directed magnetic resonance brain micro-imaging
R. Jacobs, California Institute of Technology
www.gg.caltech.edu/brain

Anatomical morphological analysis of MR brain images
D. Kennedy, Massachusetts General Hospital
neuro-www.mgh.harvard.edu/cma/ibsr

Spatially oriented database for digital brain images
S. Letovsky, Johns Hopkins University
braid.rad.jhu.edu/braid

Functional magnetic resonance imaging methods for three-dimensional mapping of brain after early injury
D. Levin, University of Chicago
mri.uchicago.edu

Enhancement of functional and neurochemical brain patterns
A. Levy, Brookhaven National Laboratory
www.ccd.bnl.gov/visualization/examples.html

A probabilistic reference for the human brain
J. Mazziotta, University of California, Los Angeles
brainmapping.loni.ucla.edu

Imaging software and methods for mapping brain development
A. Reiss, Stanford University

Spatial and temporal patterns in functional neuroimaging
D. Rottenberg, VA Medical Center, Minneapolis
pet.med.va.gov:8080/hbp

provide for a quantitative assessment of the significance of particular properties rather than a subjective impression. For this reason, integration of data into models has been designated as one of the important goals of the Human Brain Project. Computational models of neurons, running in GENESIS and in NEURON, are an integral part of several of the pilot projects (see Box 2). Computational simulations of neural processing in specific brain regions, involving both small- and large-scale neuronal networks, are also being developed (see Box 2).

Our project, termed SenseLab, uses the olfactory system as a model for addressing the problem of integrating multidisciplinary neuronal data. We describe here in more detail some of the progress that has been made, as an example of how an experimental laboratory has become involved in the development of neuroinformatics tools that are needed for our research and that also may be of broader utility.

Goals of SenseLab

Our experimental research uses a range of physiological, anatomical and molecular approaches to understanding the organization of synaptic circuits and the neural basis of sensory processing in the olfactory system[11]. In addition, we have for many years pursued parallel theoretical studies using computational models[12] to interpret the physiological data and point to testable hypotheses for further experiments.

The neuroinformatics component of SenseLab has focused on several problems that required solving in order to construct informatics tools to aid our experimental analysis of olfactory neuron properties and of odor activity maps. We have constructed these tools with a view to their more general use in assisting neuroscientists in analysing and interpreting data at the level of molecular properties, dendrites, neurons and neuronal circuits. The problems that have been addressed include: data mining; the storage and analysis of unpublished data; integrating multidisciplinary data for given neurons and parts of neurons; developing search tools to enable membrane property homologies to be identified across different neurons; automating the inputting of data into neuronal models; and ensuring data quality control. We will describe our initial progress in each of these areas.

Using unpublished data

Only a small part of neuroscience data is published; a significant problem, therefore, is how to make unpublished data accessible and useful. An example is sequence data for olfactory receptors, which could comprise up to

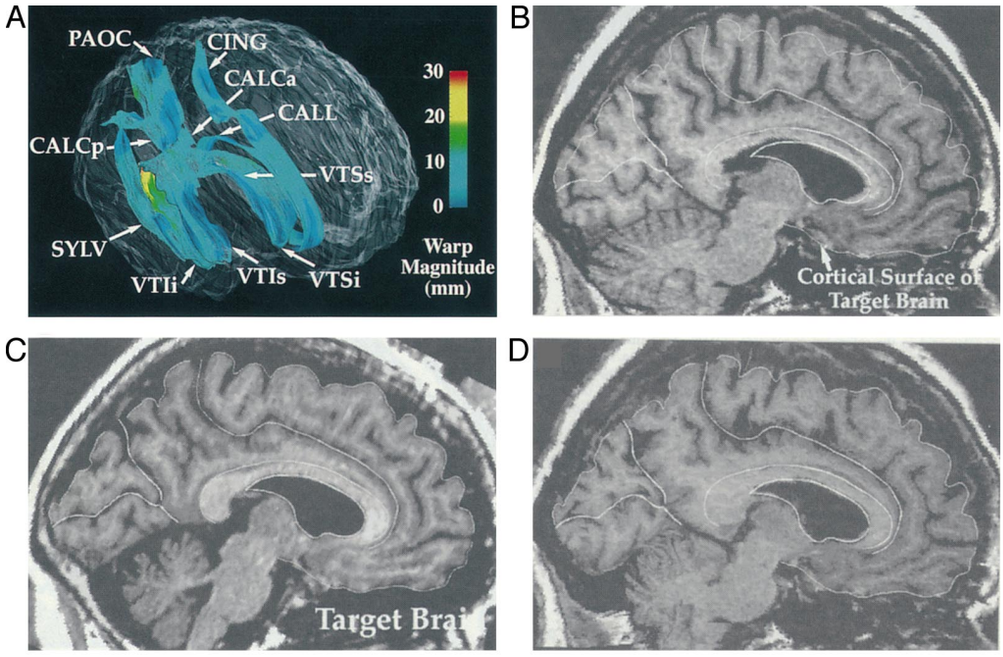

Fig. 1. ***Comparison of the magnetic resonance images of a brain of a normal subject*** **(A)** ***with an Alzheimer's patient*** **(B).** *A warping algorithm is used to bring the brain images into register* **(C)**. *The results show contraction of the cerebellar region and smaller variations in the posterior frontal and cingulate region* **(D)**. *Abbreviations: CALCa, calcerine fissure (anterior area); CALCp, calcerine fissure (posterior area); CALL, corpus callosum; CING, cingulate sulcus; PAOC, deep internal surface of the pareito-occipital sulcus; SYLV, medial surface equidistant between the banks of the sylvian fissure; VTIi, ventricular surface, inferior horn (inferior surface); VTIs, ventricular surface, inferior horn (superior surface); VTSi, ventricular surface, superior horn (inferior surface); VTSs, ventricular surface, superior horn (superior surface).* Adapted from Ref. 6 (see also loni.ucla.edu/~thompson/GORDON/1_warp_LARGE.gif).

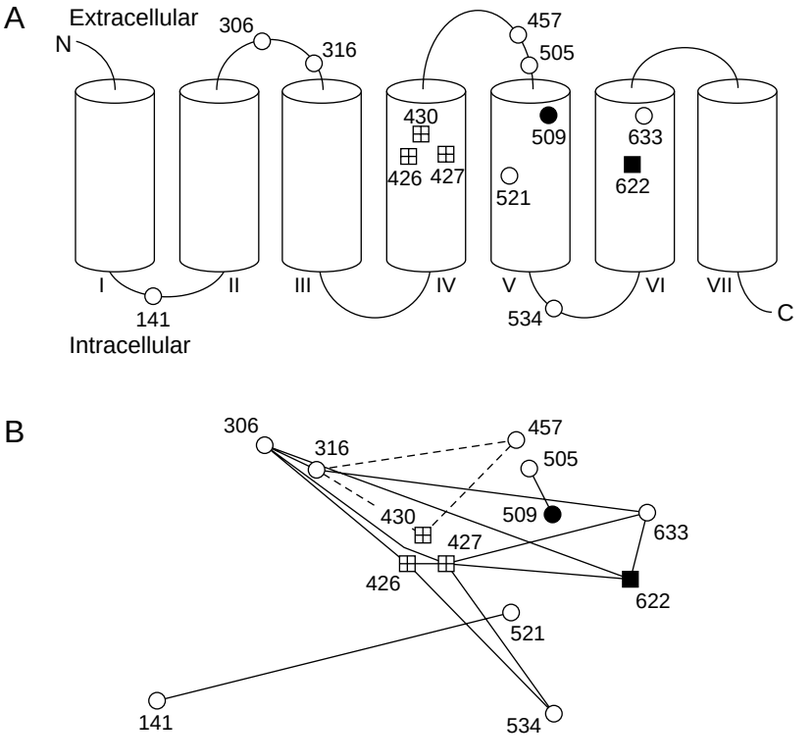

Fig. 2. ***Molecular models from data mining by the method of correlated mutation analysis (CMA).*** **(A)** *Schematic representation of the seven transmembrane helices and interhelical loops of a G-protein-coupled receptor (GPCR) belonging to the olfactory-receptor gene family. Four classes of residue positions showing correlated mutations are indicated: squares indicate residues that are identical to ligand-binding residues of a molecular model of olfactory receptor 5 (Ref. 17); filled symbols indicate residues that have been shown to bind ligands in other GPCRs; ⊞ indicate residues showing positive selection; open circles indicate residues that are unique to this CMA study.* **(B)** *Diagram of the pairwise correlations for the residue positions shown in* (A) *(Ref. 18).*

1000 genes, making them the largest family in the genome[13]. This presents an enormous challenge to laboratories attempting to clone and sequence these genes. In response to requests from several laboratories, we constructed a focused database of gene and protein sequences for these receptors, called Olfactory Receptor Database (ORDB) (senselab.med.yale.edu/ordb)[14]. A special feature is the provision of a private section in which laboratories can deposit unpublished fragments and longer sequences. A BLAST sequence comparison tool[15] enables anonymous searches for sequence similarities against a submitted sequence. If a similar, potentially homologous, sequence is found, the submitter is notified of the laboratory that deposited it, so that the laboratory can be contacted and they can discuss whether to share information and decide whether to develop a full sequence for publication. This makes use of unpublished data, and also creates a user group that enhances communication between laboratories and fosters a cooperative spirit in the field. Currently there are over 40 laboratories in the user group, and over 200 sequences in the database. Eventually one would like to have a user group that shares data freely without need for protecting unpublished data with anonymity. However, the field of olfactory receptor gene sequencing is extremely competitive and very scattered worldwide, so the private section of the database serves as a means to begin to build a community in which more open sharing can take place.

Data mining

In most neuroscience articles, published data remain locked in the hard copy journal. In contrast, sequence data, by being deposited in electronic form, are accessible to anyone for further analysis. This has given rise to a new form of research called 'data mining', in which datasets generated for one purpose can be examined for other purposes. An example is the similarity searches that led to the identification of a new form of K^+ channel[16].

An advantage of the ORDB is that it facilitates the use of published as well as unpublished data for further data mining. An example is a series of computational studies carried out on published sequences to analyse olfactory-receptor sequences for significant amino acid residue sites that might be involved in interactions with the determinants of odor molecules. One study involved construction of receptor models and automated docking of odor ligands[17]. A second[18] involved the method of correlated mutation analysis, in which pairwise analysis of residue differences through a set of sequences enables the identification of residue sites that are likely to be functionally significant (see Fig. 2). A third method involved computational analysis of positive selection moments[19]. These methods have all pointed to a consensus binding

pocket[20] which might be involved in transducing odor molecule determinants. These studies can help to guide point mutations of residues when expression systems for olfactory receptors are developed.

These studies illustrate, not only data mining, but the integrating of sequence data into molecular models in a way that can give insight into the mechanisms underlying neural processing, in this case the perception and discrimination of odors. Molecular models are a type of computational model that is new for most neuroscientists. This type of modeling will become more important as more is learned about the molecular basis of neuronal function. If the models are any good, they will suggest experiments on specific residues, which will lead back to refinements of the model, and so on. These molecular models will also eventually need to be incorporated into compartmental neuronal models (see below).

Integrating multidisciplinary data

The analysis of brain function generates multidisciplinary data ranging from protein sequences through anatomy, physiology, pharmacology, etc., to behavior. Understanding brain function requires integrating these diverse data into models for each main type of neuron; this can give insight into the mechanisms underlying the functional operations for that neuron. How to do this integration accurately and efficiently presents one of the greatest challenges to contemporary neuroscience.

The practical problem we have pursued is analysing the mitral cell of the olfactory bulb, one of the most complex cells in the nervous system. The mitral cell receives olfactory input only in its most distal dendritic tuft. It has been postulated that the EPSP response spreads through the primary dendrite to activate an action potential in the cell body–axon hillock; the action potential propagates forward into the axon as well as backward into the primary and secondary dendrites, where it activates recurrent and lateral inhibition through dendrodendritic output synapses[11,12]. The postulate that action potential initiation always occurs in the cell body–axon hillock is based on studies in other neurons[21,22]; does this hold in the mitral cell under all conditions of synaptic input? Recently we have found that the action potential can in fact be initiated in either the soma–axon hillock or the distal primary dendrite, depending on the strength of EPSPs in the distal tuft or the strength of IPSPs in the secondary dendrites[23]. These synaptic actions are in turn modulated by a variety of mechanisms, including NMDA and metabotropic glutamate receptors and centrifugal brainstem systems.

Some of these properties are illustrated by the diagram in Fig. 3. Integrating all of these data is obviously a formidable task. In order to move beyond this simple summary diagram towards a quantitative model, we have built on the compartmental approach. Based on early computer modeling studies[12], we have used the strategy of reducing the morphological complexity of dendritic trees to 'equivalent dendrites', and the complexity of a given neuron with its equivalent dendrites to a simplified 'canonical' form. In the case of the mitral cell, a minimum of compartments represents the dendritic tree: three compartments for the primary dendrite, and three compartments for the combined secondary dendrites. This is sufficient to enable the critical properties of the distal tuft, the primary dendrite and the secondary dendrites to be represented in their true distribution. To test for the generality of this approach we have applied the canonical concept to other neurons, and

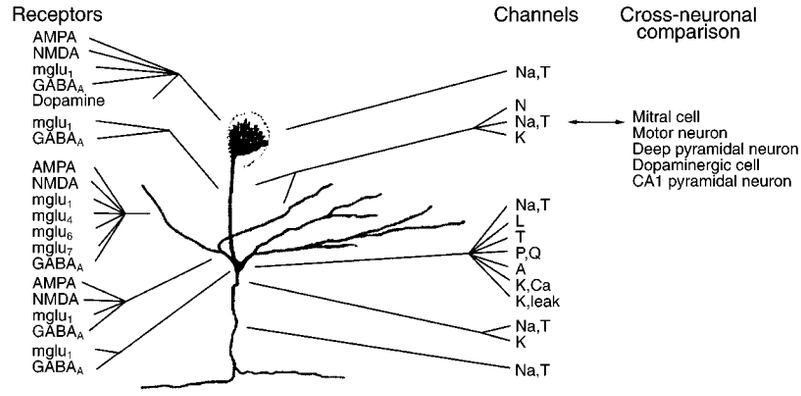

Fig. 3. ***Integration of neuronal properties in the mitral cell of the olfactory bulb.*** *The mitral cell is the first relay neuron in the pathway for the sense of smell. A wide range of properties mediates synaptic responses and action potential generation in the different parts of the cell. The figure shows a Golgi-stained mitral cell, with the distribution of neurotransmitter receptors indicated on the left and the distribution of voltage-gated membrane channels on the right. On the far right is shown a summary of the neurons in which a similarity search revealed the presence of Na, T (transient Na^+ channel) in the corresponding distal dendrites (see Fig. 4). The integration and similarity search were performed by informatics tools in NeuronDB* (senselab.med.yale.edu/neurondb). *Abbreviations: K,Ca, Ca^{2+}-activated K^+ conductance; K,leak, resting K^+ conductance.*

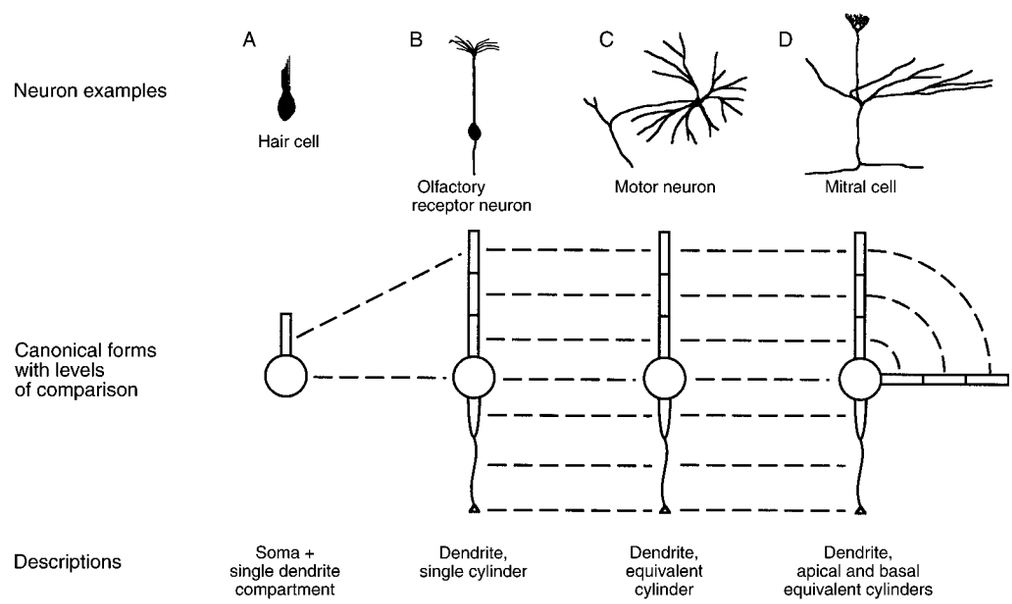

Fig. 4. ***Comparisons of neuronal properties across canonical neurons.*** *The diagram illustrates the strategy in NeuronDB* (senselab.med.yale.edu/neurondb) *of constructing reduced canonical representations of neurons with different types of dendritic trees, which enables comparisons to be made between the properties of equivalent dendritic compartments. This provides the basis for making similarity searches for neuronal properties (receptors, channels, transmitters) across different types of neurons (see Fig. 3). Using this tool, the cellular functions of different gene products can be assessed and compared in similar parts of neurons in relation to the complex matrix of properties in those parts.* From Ref. 22.

have incorporated into these canonical representations a database of neuron properties to form NeuronDB (senselab.med.yale.edu/neurondb).

For the properties of a given neuron, NeuronDB can be used in different practical ways (see Ref. 24). You can query if a given neuron has a particular type of receptor, channel or transmitter. You can query if a given compartment has a particular property. You can query for all of the receptors (or channels, or transmitters) of a given neuron, or of a given compartment, or query all properties of a given neuron or compartment. The responses come in two forms. The 'terse' form simply provides the listing of the properties queried; this is most efficient for an initial query, particularly if the list of properties is long. The 'reference' form provides three features: first, an annotation box, in which a concise narrative commentary of the studies supporting a particular property in that compartment of that neuron is entered; second, citations within the annotations for each study with hyperlinks to the abstract of that publication (in MedLine, PubMed); and, third, a combined reference list at the bottom of that entry for all of the citations, again with their hyperlinks to the abstracts. Thus, the 'terse' mode provides the overview for integrating the different properties of a compartment or neuron, whereas the 'reference' mode provides the documentation for closer analysis and integration.

The key point is that NeuronDB is not just an archive of neuronal properties; it is a tool for enabling the user to understand the significance of a molecular property within the context of other properties contributing to the functions at a particular site within a particular neuron. This is a goal, not only for neuroscientists, but also for molecular biologists studying gene function in the emerging fields of functional genomics and pharmacogenomics.

Search tools for comparing multidisciplinary data

In addition to integrating multidisciplinary data for a given neuron, we need to compare data across different neurons in order to determine similarities or differences in properties underlying the functions of different neurons. For example, in analysing the significance of distal dendritic action potential initiation, it is important to compare the properties of the mitral cell distal dendrites with those of distal dendrites of other cells in order to assess the functional significance of this type of activity. This is analogous to searching for similarities in sequence databases, except that the problem is much more difficult because neuroscience data are not one-dimensional, like sequences, but multidimensional. Can neuroinformatics solve this problem, and provide tools for searching neuroscience data that will be as indispensable for carrying out neuroscience research as the tools for searching sequence data are for carrying out research in molecular biology?

The biggest obstacle to beginning to solve this problem is the morphological complexity of most neurons, which makes it seem impossible to compare properties across equivalent parts of different neurons. As already indicated in discussing the mitral cell, in NeuronDB the complexity of dendritic trees is dealt with by constructing an 'equivalent dendrite'[12] for the basic types of trees: apical, basal, stellate, etc. (see Fig. 4). Within each compartment, the basic properties of membrane channels, neurotransmitter receptors and neurotransmitter substances are identified.

Having made the database searchable by the canonical approach, we next developed informatics tools for searching for properties across neurons[24]. The essence of these tools is that, from any compartment of any neuron in the database, the user can select a particular property (ion channel, neurotransmitter receptor or neurotransmitter) and query the database for the presence of that property in the corresponding compartment of any other neuron, or in any compartment of any other neuron. For example, in our research on action potential initiation in mitral cell dendrites (see above), a query on voltage-gated Na^+ channels in the distal primary dendrite brings up several other neurons with these channels (see Fig. 3). Variations in channel density are duly noted; links to citation databases enable the user to assess immediately the experimental data, and links to ModelDB (see below) will enable the user to assess immediately the functional consequences of differing channel densities and other properties.

Similar queries on neurotransmitter receptors enable the user to assess, for example, GABAergic inputs to cell somas compared with different levels of dendritic trees in different neurons, with citation back-up and access to model simulations. These search tools will increasingly become indispensable as neuroscience data continue growing beyond the abilities of single laboratories to keep up with work that might be crucially relevant.

Construction of neuronal models and mining of model databases

The properties of a given neuron in its different compartments are the basis for construction of compartmentalized computational models for that neuron. We are presently constructing an interface between NeuronDB and a modeling environment called Model Database (senselab.med.yale.edu/modeldb). ModelDB runs models in the simulation program NEURON (Ref. 25), and can be adapted for other programs as well. We are currently building a model for the mitral cell in order to simulate the initiation of action potentials and their control by excitatory and inhibitory synaptic responses as in the experimental study of Chen *et al*.[23] An exciting prospect is that, by the use of dual patch recordings, the parameters of the model can be much more tightly constrained than has been possible in the past with most neurons using only single-electrode recordings.

The importance of ModelDB is that each neuron model constitutes a computational tool for analysing the functional properties in parallel with experimental analysis of that neuron; critical hypotheses suggested by the experiments are tested in the models, and vice versa. This greatly enhances the insights gained from experiments, and puts the functional interpretations on a firm theoretical foundation (see Refs 26,27). The interface between NeuronDB and ModelDB will allow new or revised data in NeuronDB to be inputed automatically into ModelDB, making the process of constructing models much more efficient and user friendly. Since the models are easily accessible on the Web, it will also make possible 'model mining' – that is, the free use of the models for checking published results and exploring new functional properties of a given model.

Data quality control

In archiving neuronal properties and making them Web-accessible to integrating and search tools, a

fundamental problem is how to control for the quality of the data. Many feel that this is the main problem to be solved before any widely accessible databases should be built. Sequence databases have opted for a simple solution, in which publication in a refereed journal is the assurance of data quality. In a similar manner, it can be argued that publication should ensure the quality of neuroscience data. However, neuroscience data are often controversial. Our solution to this problem is the annotation box described above; data are accepted if they have been published, but annotations enable the significance of the data to be highlighted and any controversial aspects noted (Fig. 5). Thus, rather than being a problem to be avoided, controversies about the data are incorporated into the database so that they can be addressed by experimenters and explored by modelers. Using this strategy, unpublished data can also be included, provided it is appropriately annotated.

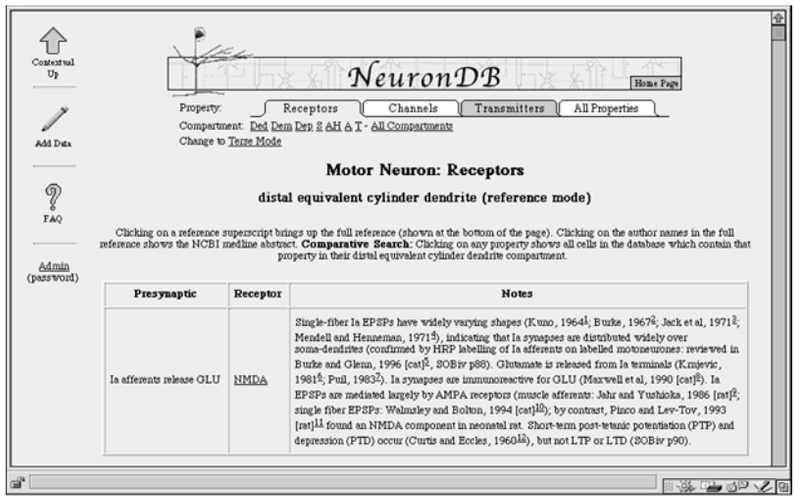

Fig. 5. ***Data quality control by use of annotation.*** *This excerpt from NeuronDB* (senselab.med.yale.edu/neurondb) *shows, for the distal dendritic compartment of the spinal motor neuron, the entry for the glutamatergic NMDA receptor. In this 'reference mode', the 'Notes' provide for annotation that gives the citation background for the development of the evidence for glutamate receptors at this level of the dendritic tree of the motor neuron. It also notes studies providing evidence that Ia afferent EPSPs are mediated largely by AMPA receptors, and contrasts this with evidence for an NMDA component, noting that the latter evidence came from the neonatal rat.*

Automating data submission

A bottleneck to future progress is the laborious process of inputting data to neuroscience databases. This process will continue to be very slow as long as neuroscience data are available only in published journal articles. Sequence data present a much more attractive model, in that journals routinely require that publication of a sequence be accompanied by submission to a sequence database, so that it is available for immediate checking and further study. This is rapidly becoming a necessity for neuroscience data. We envision that publication of research results in neuroscience journals will eventually be accompanied by a similar requirement for deposition of the essential findings in Web-accessible databases (Fig. 6). Thus, a study of the morphology of a given type of neuron, or of antibody staining for the distribution of a given channel or receptor within a neuron, or of recordings of synaptic responses or channel activity, will be accompanied by deposit at sites such as the Ion Channel Network and NeuronDB. Automating the submission process to such databases presents a challenge, but the reward will be databases and tools that are available to all users for the kinds of automated integration, search comparisons and construction of models that will greatly enhance the ability of neuroscientists to understand the neuronal basis of brain function.

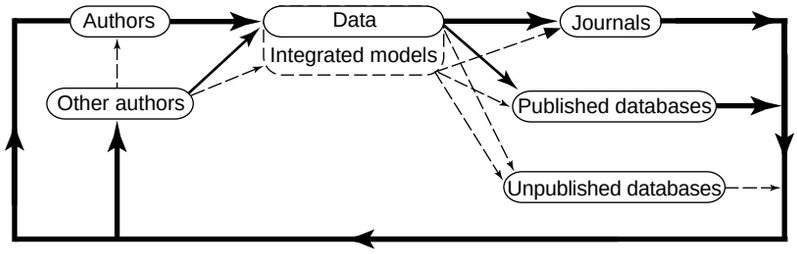

Fig. 6. ***The current (solid lines) and proposed (dashed lines) flows of neuroscience data.*** *In the current mode, authors generate experimental data that are published in journals. Journals are read by these and other authors, and the cycle repeats itself. In the case of cloning studies, submission of raw sequences to Internet databases occurs in parallel with journal publication, so that similarity searches and data mining can occur. In the proposed mode for neuroscience data, submission of data (neuronal morphology, physiology, cell images, etc.) to Internet databases will occur in parallel with journal publication, so that similarity searches and data mining can also occur with these data. In addition, when the neuroscience data are integrated into neuronal and network models, it is envisioned that journal publication will also be accompanied by deposition of the models in appropriate databases so that the models can be checked and further used. Finally, parallel submission of unpublished data to appropriate sequence, neuroscience and model databases will make much better use of grant-supported research results, leading to enhanced communication and collaboration between authors, as indicated.*

Summary and future directions

Neuroscience research is generating increasing amounts of data that go far beyond the traditional means to analyse and understand. Compared with sequence data, many types of neuroscience data, such as physiological recordings, cell imaging and brain imaging, generate huge datasets. Much of these data are lost after publication to further analysis and exchange between laboratories. The Human Brain Project is dedicated to developing a new generation of electronic methods that can make more efficient and widespread use of these kinds of data. The methods take the form of neuroinformatics tools: databases and software that can archive neuroscience data and enable these to be searched in effective ways. Once developed and proven effective, these tools will become as essential in conducting neuroscience research as are the databases and tools for gene and protein sequences. The Human Brain Project is committed to these goals by supporting pilot projects dealing with data across the spectrum of neuroscience research. The vision of the Human Brain Project is an enhancement of neuroscience research, closer cooperation between laboratories around the world, and more effective movement from research results to understanding of brain function and improvements in human health.

Acknowledgements
Our work on the Human Brain Project has been supported by the National Institute of Mental Health and the National Institute for Deafness and other Communicative Disorders (National Institutes of Health), and the National Aeronautics and Space Agency.

Selected references

1 **Huerta, M.F., Koslow, S.H. and Leshner, A.I.** (1993) *Trends Neurosci.* 16, 436–438
2 **Gobel, U. *et al.*** (1994) *Proteins* 18, 309–317
3 **Pechura, C. and Martin, J.** (1991) *Mapping the Brain and its Functions. Integrating Enabling Technologies into Neuroscience Research*, National Academy Press
4 **Roberts, L.** (1993) *Science* 252, 1794
5 **Cotton, P.** (1993) *J. Am. Med. Assoc.* 269, 1357

6 **Thompson, P.M. and Toga, A.W.** (1996) *IEEE Trans. Med. Imaging* 15, 402–417
7 **Bloom, F.E., Young, W.G. and Kim, Y.K.** (1989) *Brain Browser*, Academic Press
8 **Swanson, L.W.** (1995) *Trends Neurosci.* 18, 471–474
9 **Brinkley, J.F. and Rosse, C.** (1997) *J. Am. Med. Informatics Assoc.* 4, 165–183
10 **Van Essen, D.C. and Drury, H.A.** (1997) *J. Neurosci.* 17, 7079–7102
11 **Shepherd, G.M. and Greer, C.A.** (1998) in *The Synaptic Organization of the Brain* (Shepherd, G.M., ed.), pp. 159–203, Oxford University Press
12 **Rall, W. and Shepherd, G.M.** (1968) *J. Neurophysiol.* 31, 884–915
13 **Buck, L.D. and Axel, R.** (1991) *Cell* 65, 175–187
14 **Healy, M.D. *et al.*** (1997) *Chem. Senses* 22, 321–326
15 **Altschul, S.F. *et al.*** (1990) *J. Mol. Biol.* 215, 403–410
16 **Ketchum, K.A. *et al.*** (1995) *Nature* 376, 690–695
17 **Singer, M.S. and Shepherd, G.M.** (1994) *NeuroReport* 5, 1297–300
18 **Singer, M.S. *et al.*** (1995) *Recept. Channels* 3, 89–95
19 **Singer, M.S. and Shepherd, G.M.** (1996) *Recept. Channels* 4, 141–147
20 **Shepherd, G.M., Singer, M.S. and Greer, C.A.** (1996) *Neuroscientist* 2, 262–271
21 **Eccles, J.C., ed.** (1958) *The Physiology of Nerve Cells*, Johns Hopkins Press
22 **Stuart, G. *et al.*** (1997) *Trends Neurosci.* 20, 125–131
23 **Chen, W.R., Midtgaard, J. and Shepherd, G.M.** (1997) *Science* 278, 463–467
24 **Mirsky, J.S. *et al.*** (1998) *J. Neurosci. Methods* 82, 105–121
25 **Hines, M.L. and Carnevale, N.T.** (1997) *Neural Comput.* 9, 1179–1209
26 **Segev, I., Rinzel, J. and Shepherd, G.M., eds** (1995) *The Theoretical Foundation of Dendritic Function*, MIT Press
27 **Koch, C. and Segev, I., eds** (1998) *Methods in Neuronal Modeling: From Synapses to Networks*, MIT Press

LETTERS TO THE EDITOR

The pathophysiological mechanism underlying Rasmussen's encephalitis: a debate

Rasmussen's encephalitis (RE) is a pediatric syndrome characterized by epileptic seizures associated with unilateral inflammatory process in the brain[1,2]. It is, according to several recent lines of evidence, an autoimmune disease initiated by circulating autoantibodies that, under certain circumstances, gain access to the brain where they interact with the ionotropic glutamate receptor 3 (iGluR3) subunit[3–5]. However, the crucial events triggered by these autoantibodies are currently under debate. In a recent article in this journal[6,7], I briefly described the findings of Twyman *et al.*[4], who proposed that the pathophysiological mechanism underlying RE involves excessive activation of iGluR3-containing ion channels by these autoantibodies and subsequent excitotoxic neuronal cell death. This type of cell death has been implicated in several brain disorders and chronic neurodegenerative diseases[8]. By contrast, He *et al.*[9] have recently suggested that neuronal cell death in RE patients might result from the activation of the complement system[10,11] by the anti-iGluR3 autoantibodies. In any event, both types of mechanism could potentially induce recurrent seizures due to uncontrolled activation of ion channels (primarily, by the autoantibodies or by glutamate leaking from damaged cells or both), as well as recruitment of various components involved in inflammation to the site of injured neurons, thus giving rise to the symptoms, clinical features and histopathology typical of RE (Refs 1–3,12).

The hypothesis of excitotoxicity-mediated neuronal cell death in RE is mainly based on the following observations:

(1) Following the immunization of rabbits with a large portion of the extracellular N-terminal region of iGluR3 (fused to bacterial trpE protein), the filtered sera but also IgG fractions of RE-like symptomatic rabbits elicited rapid, reversible, and voltage-independent opening of cationic channels in kainate (KA)-responsive cultured neurons[4].

(2) Similar currents were also elicited by filtered sera and IgG fractions of RE patients[4].

(3) The final passage filtrate of these sera as well as control sera depleted of low molecular weight substances by repetitive filtration (as carried out for the sera of symptomatic rabbits and RE patients) did not elicit currents in KA-responsive cultured neurons[4]. These controls included sera of pre-immune rabbits, sera containing antibodies raised against a similar portion of iGluR5 and antibodies against the β2 subunit of the neuronal nicotinic acetylcholine receptor, and sera sampled from healthy individuals and patients with other neurological diseases. Hence, the electrophysiological responses could not be attributed to traces of glutamate or other putative low molecular weight agonists in the sera containing the anti-iGluR3 antibodies.

(4) The currents elicited by sera or IgG fractions were inhibited by CNQX, a competitive antagonist of AMPA and KA receptors, as well as by a synthetic peptide corresponding to the suspected antigenic epitope (residues 372–395 of iGluR3). These inhibitory effects were specific (see Ref. 4 for controls) and voltage-independent, indicating that CNQX and the synthetic peptide do not directly block the channel pore but probably interact with other sites (CNQX with residues located in the agonist-binding pocket and the synthetic peptide with the antibody paratope).

(5) Channel-activating sera of rabbits and RE patients labelled iGluR3-transfected kidney cells but not cells transfected with other iGluR subunits[4].

It is puzzling that He *et al.*[9] could not replicate these electrophysiological recordings, despite the application of plasma 'filtrates' or IgG fractions (of RE-like symptomatic rabbits) that contained high titres of anti-iGluR3 antibodies raised against a similar antigen that consisted of a fusion protein of glutathione-S-transferase–iGluR3 (residues 246–455)[9] versus TrpE–iGluR3 (residues 246–458)[3,4]. These experiments were performed in KA-responsive cultured neurons that contained the iGluR3 subunit. No currents were detected by He *et al.*[9] even in the presence of cyclothiazide, a compound that considerably attenuates desensitization of AMPA receptors[13].

The major observations that support the idea that RE is mediated by complement activation are:

(1) Plasma or serum 'filtrates' of RE-like symptomatic rabbits induce death of cultured cortical neurons in a concentration-dependent manner[9]. This cell death was neither inhibited by CNQX nor by a non-competitive AMPA receptor antagonist (GYKI 52466), whereas in control experiments these antagonists did inhibit AMPA-induced cell death[9].

(2) Neither IgG fractions alone nor IgG-depleted plasma of symptomatic rabbits alone induce cell death, but when mixed together they reconstituted the cytotoxic capability[9]. However, IgG-depleted plasma which was preincubated at 56°C for 30 min in order to heat-inactivate the complement system failed to reconstitute the cytotoxic capability[9].