Neuroinformatics research at Yale University as part of the national Human Brain Project (HBP) is exploring the integration of multidisciplinary sensory data, using the olfactory system as a model domain. The overall name for the project is SenseLab. SenseLab’s neuroinformatics activities fall into three main areas:

I Building databases and related informatics tools to support olfactory research, with a particular focus on databases that support our experimental work at Yale and can also serve as resources for the field as a whole. These currently include databases containing a) sequence data for olfactory receptors, b) data on odor molecules, c) properties of neurons as a whole, d) properties of neurons in their specific dendritic compartments, and e) computer models of neurons and their dendritic compartments.

I Using computational methods to analyze and integrate data being collected experimentally and to help guide further laboratory experiments. These methods include a) computational analysis of olfactory receptor sequences, b) construction of molecular models of olfactory receptors and of their binding to odor molecules, and c) construction of models of neurons and their dendritic compartments.

I Informatics research that uses these activities as a context to develop new informatics technologies. We have developed a flexible data model (which we have named EAV/CR, entity-attribute-value with classes and relationships) designed to facilitate the storage and integration of diverse heterogeneous data within a single unifying framework.

Affiliation of the authors: Yale University, New Haven, Connecticut.

This work was supported in part by grant R01 DC03972 from the National Institutes of Health and grant G08 LM05583 from the National Library of Medicine.

Correspondence and reprint requests: Perry L. Miller, MD, PhD, Center for Medical Informatics, Yale University School of Medicine, P.O. Box 208009, New Haven, CT 06520-8009; email: perry.miller@yale.edu.

Received for publication: 6/19/00; accepted for publication: 9/8/00.
without requiring changes to the database schema or to the routines that access it. Our five main HBP databases are currently operational within a single integrated EAV/CR framework.

This paper describes these three neuroinformatics themes that underlie our HBP work. In addition, we believe that these activities, viewed as a whole, provide a useful pilot model that illustrates the national HBP approach. As described in more detail in the Discussion section, SenseLab demonstrates how a HBP project can include 1) data repositories of well-defined types of experimental results at the gene, protein, cell, and circuit level, 2) tools that analyze and integrate that data, including computer-based models, and 3) basic informatics research to develop more sophisticated tools for dealing with the complexity of neuroscience data.

Figure 1 helps place the integration of multidisciplinary sensory data into a broad perspective. The figure outlines the major levels at which processing occurs in a sensory system, starting at the genetic level, moving up through the synaptic, neuronal, and brain-pathway levels, and ultimately reaching the behavioral level. To understand the overall behavior of the olfactory system, experimental research must be performed at each of these levels and then analyzed in an integrated fashion. At each level, a range of different phenomena are investigated using a variety of different experimental techniques.

Historically, laboratories have tended to focus their research at only one or two of these levels, in part
because the laboratory techniques used at each level are very different. In addition, the data have typically been analyzed primarily at the level at which they were gathered.

The storage of increasingly large amounts of data in computer-based form, however, provides an opportunity to perform integrated analysis in a way that was previously not possible. A central challenge in the field of neuroinformatics concerns how to store these many diverse types of data, how to link related data among the different levels, and how to perform integrated analyses. Our work on the olfactory system represents one step in starting to explore these issues.

Background

This section describes the background of the SenseLab project from several perspectives: 1) the experimental research on the olfactory system on which our neuroinformatics activities are based, 2) the use of computer modeling at the molecular and the neuronal levels to support that research, 3) certain basic informatics research challenges that this domain presents, and 4) the relationship of this project to other HBP research.

Experimental Olfactory Research

The olfactory system is the subject of increasing experimental and theoretic interest because of several developments; specifically, the discovery of a large gene family of putative odor receptors\(^2\); the convergence of a number of experimental studies, providing a consensus on the principles underlying the functional organization of the system\(^3\)–\(^5\); and the effective use of molecular, neuronal, and circuit models for simulating and testing functional operations of neural circuits.\(^6\)–\(^8\)

Identification of olfactory receptors began with the cloning and sequencing of a large gene family of G-protein coupled receptors.\(^2\) These receptors have been shown to respond differentially to odor molecules\(^9\)–\(^12\). Understanding of their function has been frustrated, however, by difficulties in coordinating efforts to clone and sequence this extremely large gene family (estimated to contain 1,000 members), and difficulties in experimentally expressing the receptors in heterologous systems. A related problem is the difficulty of knowing whether these receptors are being adequately tested for their preferred ligands (odor molecules).

Molecular and Neuronal Computer Modeling of the Olfactory System

To aid in the efforts to clone and analyze these receptors, computational analysis of the receptors has been carried out to gain evidence for the possible nature of odor ligand–odor receptor interactions. This has included development of a new method of sequence analysis called correlated mutation analysis, and construction of molecular models of receptor-odor interactions. These studies have been the first to apply molecular modeling techniques to the analysis of odor ligand-odor receptors.\(^13\)–\(^17\)

At the level of the whole neuron, integration of data regarding membrane channels, membrane receptors, and secretion of neurotransmitters represents a daunting challenge. Use of compartmental models has become established as a leading method for achieving this integration in a rigorous manner. Using the olfactory mitral cell as a model, we have carried out precise simulations of complex nonlinear behavior in the dendrites that underlie key steps in information processing at this stage.\(^18\)

Underlying Informatics Research Challenges

The olfactory system, like other parts of the brain,\(^19\),\(^20\) is characterized by complexity and heterogeneity. In the typical experimental neuroscience setting, the researcher sets out to study an aspect of particular cells or a particular part of the nervous system. The information that is captured through experiments is then stored in a database for future analysis.

One fundamental informatics research question concerns how this heterogeneous, interrelated data should best be structured so that it can be searched and manipulated in a robust, integrated fashion. In practice, an experimental “database” may be just a collection of text files or spreadsheets. Alternatively, a project may use a database management system (DBMS). The use of a DBMS does not, however, solve the problem of integration. The bioinformatician must provide a model for representing and organizing the different classes of data to be stored.

Also, as scientific knowledge evolves, especially with a pooled database that stores highly heterogeneous data, it becomes desirable to include metadata; that is, higher-level data describing the structure of the neuroscience data. The goal is to help insulate the database design as well as the analysis routines, including the user interface code, from the contents of the database as the data increase in complexity and
diversity. As described later in this paper, we have developed an approach for dealing with these problems that we call the EAV/CR data model.

Relationship to Other Work in the National Human Brain Project

The national Human Brain Project supports a wide range of endeavors at many different levels of neuroscience research. A good survey of this work is available at the HBP Web site, http://www.nimh.nih.gov/neuroinformatics/index.cfm. As outlined there in detail, HBP work includes behavioral research, whole-brain imaging and mapping for diverse purposes, neurophysiologic recording, structural studies at the cellular and fine structural level, and computer-based modeling at several levels. Our work on SenseLab involves developing a closely integrated set of databases and software tools, including computer-based molecular and neuronal modeling, that span a considerable range of these levels of research within a single project. As a result, SenseLab provides a pilot model of the type of multidisciplinary and multilevel integration that the HBP is working to achieve.

Building Databases and Informatics Tools to Support Olfactory Research

The first area of our neuroinformatics activities involves developing databases and related tools to support olfactory research. This section describes the five databases that we have built to date, which are currently in different stages of maturity.

- **ORDB** (Olfactory Receptor Database) contains information about olfactory receptors. ORDB has been an operational resource supporting the field for several years.

- **OdorDB** is a prototype database containing information about ligands (odor molecules). This database is still in the early stages of development and refinement.

- **CellPropDB** is a recently developed database containing data on the membrane properties of different types of neuron.

- **NeuronDB** is a database containing data about neuronal properties in different neuronal compartments. NeuronDB is currently populated with information describing the olfactory mitral cell and a range of neuron types in other brain regions and has been available as a pilot resource for the field for over a year.

- **ModelDB** is a database whose structure closely parallels that of NeuronDB. It was designed to store computer-based neuronal models and to facilitate the viewing, searching, uploading, and downloading of those models via the Web. ModelDB is currently undergoing active refinement through use at Yale and at one closely collaborating institution.

**ORDB: An Olfactory Receptor Database**

Olfactory receptor (OR) genes, which are estimated to comprise up to 1,000 subtypes per species, may be the largest gene family in the genome. Cloning and sequencing this huge family and finding the preferred ligands for each gene product is a more daunting task than that attempted for any gene family previously in neuroscience or in the human genome in general. In response to requests from sequencing laboratories for assistance in this endeavor, we developed ORDB21,22 (http://ycmi-hbp.med.yale.edu/senselab/ordb), a Web-accessible database for organizing the information about these receptors and receptor genes and providing informatics tools for searching the database for sequence similarities and insights into the structure and function of these receptors.

Investigators also asked that this database include unpublished sequences, enabling searches for possible duplications of whole or partial OR sequences. The goal was to facilitate coordination of sequencing efforts among the different laboratories. In response to this request, we developed one innovative aspect of ORDB—its ability to store both private and public OR sequences.

Public OR sequences include those that have been published in the literature or are found in publicly accessible databases such as GenBank. Access to public OR sequences is free via the Web. Private OR sequences are contributed by participating laboratories and are not accessible to the public. Only the contributing laboratory is allowed to view the details describing a private OR sequence. Other participating laboratories, however, are able to perform BLAST sequence-comparison searches against all public and private sequences. If a match is found against a private sequence from another laboratory, the user must contact that laboratory to obtain detailed information about the matching sequence.

The significance of ORDB includes the following: It was built in response to a request from the field and addresses a clearly perceived problem in the field. It has supported the development of the first molecular
models of olfactory receptor-odor molecule interactions. In addition, it is one of the first databases that allows controlled access to unpublished neuroscience data in addition to public data.

OdorDB: A Prototype Database of Odor Molecules

To aid investigators in testing expressed receptors, we are currently developing and refining OdorDB, a database that contains information about odor molecules. When fully operational, OdorDB will have tools that allow it to be searched by chemical type (e.g., alcohols, aliphatic acids), numbers of carbon atoms, cyclic components, odor sensation, and such. It will contain structures of odor molecules in the form of image maps. We will also develop pointers in the database engine to the image files stored separately. A future goal is to link specific odor molecules described in OdorDB to specific ORs stored in ORDB in a way that records experimental data relating the behavior of a given OR when presented with a given odor molecule.

CellPropDB: Cell Properties Database

A critical problem in neuroscience is the need to identify the genes and gene products (proteins) expressed by each cell type. This problem will become increasingly important with the completion of the sequencing of the human genome. Already there are initiatives to identify the genes associated with different regions of the brain. This information will have no functional significance, however, until it is related to the cell types in each region.

CellPropDB is a recently developed database that addresses this problem by providing a repository for information describing three types of membrane properties in a selected subset of projection neurons in different brain regions. These properties are the different types of synaptic receptors, ion channels, and neurotransmitters found in each cell. For CellPropDB to be practical, we foresee that automated tools will be needed to help populate the database from sources such as PubMed, so that each property can be supported by an accumulation of citations from the relevant literature. The development of such tools is a challenging problem that we are currently working to solve. CellPropDB will also serve as an interface between the literature and NeuronDB (described below), where the properties are further distributed among the compartments within each neuron, as this more detailed information becomes available.

NeuronDB: An Integrative Database of Neuronal Properties

Understanding the significance of a specific neuronal membrane property requires relating that property to the integrative context of the neuron as a whole, and comparing it with other neurons. This process is
hampered by the sheer volume of data and by the fact that the data are produced by individual laboratories without a common structured environment for storing, retrieving, and integrating those data.

To deal with this problem in our own research on olfactory circuits, the development of an organized database of neuronal properties correlated with neuronal morphology became crucial. We call this database NeuronDB. 23, 24 Not only can NeuronDB enhance our experimental analysis of neuronal properties, it can also provide efficient methods for the construction of neuronal models. An ultimate goal is to develop automated procedures for rapidly, accurately, and efficiently constructing models of realistic neuronal circuits. None of the current archives of neuronal properties, either in print or electronic formats, provides these capabilities. (See, for example, the Ion Channel Network, at www.le.ac.uk/csn.)

Our experience with compartmental modeling of neurons suggested a strategy of constructing a database provided with tools for integrating neuronal properties into neuronal representations that could be directly input into models, and making this database searchable across different neurons. This provides for the first time a database with search tools for neuronal properties equivalent to the search tools that have made sequence databases essential for molecular biology.

Three critical organizational variables define the main dimensions of the database: neurons, properties and compartments. NeuronDB facilitates data integration within a neuron in a variety of ways. The simplest form allows the user to ask for all properties (membrane channels, synaptic receptors, neurotransmitters) within all compartments of a given type of neuron. This gives multiple pages of data, systematically organized by compartment (e.g., distal apical dendrite, middle apical dendrite, proximal apical dendrite, soma, axon hillock, axon, and axon terminal), repeated for each of the three types of properties. This easily reviewed summary provides an integration of the different data that will serve many purposes. We have tested the value of this integrative function first in supporting the construction of a compartmental model of the mitral cell, which we have used to simulate both back and forward propagation of the impulse in the primary dendrite of the mitral cell. 18, 25

The database can also facilitate data integration across neurons in several novel ways. The key to comparing data across neurons is provided by the concept of a canonical neuron (Figures 2 and 3). A canonical form of a neuron uses the least number of compartments sufficient to represent essential neuron functions. 26 Since some neuronal types have very different morphologies, we have created several corresponding canonical neuron types. The database enables the properties of neurons of different morphologies to be compared by identifying equivalent compartments in canonical representations of different neurons. For...
**Figure 4** Two screens that provide access to data in ModelDB, by region (left) or by neurotransmitter (right).

**Figure 5** ModelDB screen listing all models that contain the neurotransmitter Gaba.

**Figure 6** ModelDB screen displaying information that describes a specific model.
NeuronDB we have started with the three main types of projection neurons in the olfactory pathway—olfactory receptor neuron, mitral/tufted cell, and cortical pyramidal neuron. We are extending this to the projection neurons of several other brain regions as well.

ModelDB: A Database of Computer-Based Neuronal Models and their Results

ModelDB is a database built to facilitate Web-based access to neuronal models stored in a structured fashion that closely parallels the structure of NeuronDB. A major focus of our olfactory work is to use the experimental data on membrane properties to construct realistic compartmental models of soma and dendrites. These models allow one to test hypotheses of neuronal function in a rigorous manner. A drawback of these types of models so far is that it is usually not possible for other laboratories to test a given model. ModelDB is a Web-accessible database (http://ycmi-hbp.med.yale.edu/senselab/modeldb) designed to help to solve this problem by permitting and encouraging widespread testing, validation, and enhancement of neuronal models, in the olfactory system and in other systems.

ModelDB is in pilot operation, being used primarily by a single user (M.H.) at Yale. It currently contains seven models of three types of cells—mitral cells, thalamic relay neurons, and thalamic reticular neurons. The database can be accessed in several ways, including by region or by neurotransmitter, both illustrated in Figure 4. For example, on the small screen (Figure 4, right), clicking on “Gaba” yields a list of models that contain this neurotransmitter, as shown in Figure 5. Clicking on a specific model name (e.g., “DLGN”) on the screen shown there yields a screen describing that model (Figure 6). The column on the left of the screen shown in Figure 6 contains links to the various components of that model (represented using the NEURON simulation language). These components can be inspected individually. Alternatively, the model as a whole can be downloaded and run.

In addition to serving as a Web-based repository for neuronal models, ModelDB also has the potential to serve as a vehicle for helping record and organize a sequence of neuron modeling simulations conducted over a period of time. To explore how the process of neuronal modeling might be assisted in this fashion, pilot work was performed several years ago to build such an environment.27 This early version of ModelDB was designed to support the iterative process of neuronal modeling by storing the incremental models and their results in an organized fashion. This approach also structured the various components of a model (e.g., into parameters unlikely to be varied and parameters likely to be varied) to facilitate the organized maintenance of a set of modifications to a model over time. In the future, we plan to incorporate such a database approach to supporting the iterative modeling process into the current ModelDB framework.

Using Computer Models to Help Understand Experimental Data and Guide Future Experiments

Using these databases, several projects have been carried out using computational methods to give insight into key questions regarding information processing in the olfactory pathway. This section describes examples of this work.

New Methods of Computational Sequence Analysis to Identify Functionally Significant Amino Acid Residues

No crystal structure is available for olfactory receptors, and given the difficulties of crystallizing membrane proteins, one is not likely to be forthcoming. In the absence of a crystal structure, computational approaches have been valuable in generating specific hypotheses about olfactory receptor function. A recently developed method to locate potential functional sites is correlated mutation analysis.28 Correlated mutation analysis uses multiple protein sequences to identify functional residues without the intermediate step of structural information. The method compares several members of a protein family, such as different olfactory receptor subtypes, and scans for pairs of residues that remain constant or mutate in tandem. Figure 7 illustrates the method.

The correlated mutation behavior shown by these residues indicates an increased likelihood that they are functionally or structurally interdependent. The nature of this interdependence can be predicted from the locations of the correlated residues, whether in the transmembrane domains, extracellular loops, or intracellular loops. When we applied correlated mutation analysis to rat olfactory receptor sequences, the results pointed to an odor-binding pocket similar to the epinephrine-binding pocket of the beta-adrenergic receptor.29 Variations in the binding pocket residues are believed to account for the different odor responses seen in different olfactory receptor subtypes. Molecular models, explained below, provide further support for the binding pocket.
Molecular Modeling of Olfactory Receptor-Odor Molecule Interactions

The recent availability of expression systems has made it possible to test the odor responses of olfactory receptors (ORs) and probe the effects of amino acid mutations.9,10 There is now a critical need to understand these interactions at the atomic level. We believe this need can be addressed by the same approach used for the beta-adrenergic studies, site-directed mutagenesis, and molecular modeling.

We have built molecular models of olfactory receptors and automatically docked odor ligand molecules into these models. The aim was to identify the potential odor-binding pocket by methods independent of correlated mutation analysis and, further, to predict how amino acids in that pocket interact with functional groups on the odor ligand. We also want to predict the relative affinities of different odor ligands and the effects of amino acid substitutions on odor ligand binding.

Based on an earlier model of the rat OR5 receptor,13 we carried out a pilot study on the rat I7 receptor, which has been shown to respond preferentially to the aldehyde n-octanal.9,10 We identified transmembrane domains by multi-sequence hydrophobicity profiles, packed the helices based on the 7.5 Å resolution density map of rhodopsin (a related G-protein coupled receptor), and refined the structure with energy minimization and molecular dynamics.

Octanal and related ligands were automatically docked in the I7 receptor model (Figure 8). The binding pocket matched that predicted by the correlated mutation analysis. The most important residue in this pocket was lysine 164, which formed an electrostatic interaction with the carbonyl group of octanal. Notably, this residue position corresponds to the histidine previously predicted to bind odor molecules in the rat OR5 and other ORs.13,28 The lysine was stabilized by a nearby aspartate (204) in the receptor. Five hydrophobic or neutral residues formed contacts with octanal and are expected to confer specificity for the alkyl chain.

We also docked several octanal analogs into the model. Predicted affinities for the aldehyde series compared favorably with experimental results. We next carried out residue substitutions equivalent to site-directed mutagenesis. Alanine substitution of lysine 164 substantially reduced I7 affinity for octanal, whereas substitution of aspartate 204 enhanced affinity. These results provide leads for site-directed mutagenesis experiments.

Compartmental Modeling of Action Potential Generation in Neuronal Dendrites

Computer modeling of mitral cell electrical properties was motivated by our desire to understand the shifting site of action potential (AP) initiation with stimulus location and intensity, which we had shown in previous experimental studies.25 These kinds of electro-
Physiological experiments yield data that reflect a complex consequence of cell morphology (electrical cable properties) and a rich and spatially varying set of ion-selective, voltage-gated membrane channels. Computer models that embody our qualitative intuitions about the mechanisms that are implied by this data provide rigorous consistency across our variety of stimulation protocols and help us design new experiments that test more directly the key notions in our model. (Figure 9 illustrates this type of modeling.)

Drawing on data contained in NeuronDB, a computer model of the mitral cell was constructed, which demonstrated that AP initiation site as a function of electrode current stimulation strength could be quantitatively explained by 1) lower AP threshold in the

**Figure 8** Details from the 17 olfactory receptor molecular model, showing octanal (rendered as a surface) in the predicted odor-binding pocket (rendered as sticks). White indicates carbon atoms, dark gray indicates oxygen, and the arrow indicates nitrogen. Medium gray indicates the carbon atoms of octanal. This model highlights a critical interaction between lysine 164 and the carbonyl (double-bonded oxygen) atom of octanal.

**Figure 9** The results of two experiments (A and B), showing how neuronal modeling of the olfactory bulb mitral cell can help us understand experimental results. The dashed lines show recordings of membrane potential made at two locations (d and s). The solid lines show the predictions of our neuronal model. The goal of this work is to produce a model that matches the data through the rising phase of the spikes. In experiment A (left), location d is stimulated with a low current, whereas in experiment B (right), location d is stimulated with a high current. Intuitively, one would expect that the resulting action potential would start at site d and move toward site s, which is seen in experiment B with the high-current stimulus. In experiment A (with the low-current stimulus), however, the spike somehow manages to start at site s and propagate back out to site d (where the actual stimulation occurred). The neuronal model is able to explain this shift in the location of spike initiation in terms of the inhomogeneity of threshold and of voltage gradients along the dendrite. (From Shen et al. Used with permission.)
axon initial segment and 2) the membrane voltage gradient along the primary dendrite. Current studies indicate that the same model is consistent with experiments which showed double APs and greater shifts in AP timing with synaptic stimulation, as shown by Chen et al.

Basic Neuroinformatics Research and Development: The EAV/CR Data Model

This section describes the third theme of our neuroinformatics work, in which we are performing basic informatics research in the context of the two neuroinformatics themes described above and folding the results of the research back into our overall approach. This section discusses one major focus of our basic informatics research, the development of the EAV/CR data model, which represents a potentially powerful and flexible approach to the representation of heterogeneous bioscience data.

Overview of the EAV/CR Data Model

Like the central nervous system as a whole, the olfactory system is characterized by its complexity and heterogeneity. This presents major challenges in attempts to store and manipulate neuroscience data. Many different types of data will need to be stored, and the nature of each type of data (e.g., diverse experimental results) will evolve over time. To deal with these complexities using conventional (relational) database technology requires constant extension and modification of the database tables. This in turn requires that the programs (such as those that perform database queries and implement the user interfaces) be constantly rewritten. We have developed an approach that avoids these major difficulties.

Our approach is based on a data model that we call EAV/CR (entity-attribute-value with classes and relationships). The EAV/CR approach allows diverse types of data to be accommodated within a single unifying data model, as described below. A key feature involves including a data library containing metadata (data describing the data elements) that can be used to guide the manipulation of the data. When new types of data are included in an EAV/CR database, a description of that data is added to the metadata. The programs that manipulate the data consult this metadata when formulating database queries and when presenting a user interface. As a result, using the newly defined metadata, these programs can operate on new types of data without modification.

In this section, we first describe the basic EAV data model and then describe how we have extended this model to develop the EAV/CR approach.

The EAV Data Model

In the EAV data model, data element names (such as “sequence,” “species,” “channel type,” “receptor type,” and “pH”) are not hardwired into the database as table column headers, as they are in a traditional relational database. Rather, they are stored as data. In addition, metadata describing each data element are stored in a data library, where the data item definitions can be readily created, viewed, and edited by the user.

Conceptually, an EAV design involves a single table with three columns, an entity (such as an olfactory receptor ID), an attribute (such as “species,” which is actually a pointer to the metadata table), and a value for the attribute (e.g., “rat”). (In practice the design is somewhat more complicated than this simple conceptual model, for a variety of pragmatic and efficiency-related reasons.) The EAV table has one row for each attribute-value pair, which represents one “fact” stored in the database. The EAV design has been widely used in several well-known electronic patient record systems, including the pioneering HELP system and the Columbia-Presbyterian Clinical Data Repository. The Yale Center for Medical Informatics has used the EAV approach in several clinical databases, including Trial/DB (formerly called ACT/DB), a clinical trials database.

The EAV/CR Data Model

To handle the complexity of neuroscience data, the EAV/CR data model has extended the basic EAV data model in two major ways:

- **Classes.** In the EAV data model, data elements can have only simple values. For example, a value could be an integer, a floating point number, or a string. To represent the complexity of neuroscience data, it is important that a value can be a complex object, or “class,” as described below in more detail. For example, a value might be a specific channel with several properties, such as the ion involved, the “rate” (e.g., “fast” vs. “slow”), and the result (e.g., hyperpolarization vs. depolarization). One cannot represent such a complex object as a value with the simple EAV data model.

- **Relationships.** In the EAV data model, there is no way to explicitly represent the biological relationships between items in the database. An example of such a relationship is “modulation,” which
could relate a channel, a receptor, and an effect. The EAV data model does not accommodate the representation of such relationships.

The EAV/CR approach extends the EAV data model to include these capabilities. To help make the approach more concrete, Figure 10 gives a simplified example of an EAV database and an EAV/CR database.

- In the EAV database (Figure 10, left), each row corresponds to a single fact. In this example, we show two facts, both involving a specific olfactory receptor with the name “or_1.” The two facts indicate the species and sequence associated with or_1.

- In the EAV/CR database (Figure 10, right), we see how a relationship involving complex values can be stored.

To understand this example, we first need to discuss metadata. The metadata for this simplified example would include the following information:

- **receptor_channel_association**
  - [attributes: channel, receptor]

- **channel**
  - [attributes: sequence, ion]

- **receptor**
  - [attributes: agonist, sequence]

(The actual metadata includes more attributes for these items and additional information about the nature of each data item and each attribute.)

The simplified metadata shown above lists the attributes of each complex data type; in this case, the relationship “receptor_channel_association” and the classes “channel” and “receptor.” (Notice that an attribute may itself be the name of a complex data type.) In the simplified EAV/CR database, we see how a specific receptor_channel_association (“rca_1”) is defined, which involves a specific channel (“chan_25”) and a specific receptor (“rec_33”), each of which is a class that has its properties stored in the database. In contrast to the EAV database, where each row represents a separate fact, in the EAV/CR database all the facts shown in this example are tied together into a relationship (rca_1), which relates two classes (chan_25 and rec_33), whose properties are also recorded in the database.

The actual schemas for our current EAV/CR metadata and data have been described in greater detail by Nadkarni et al.30

**Integrating Our Five Current Human Brain Project Databases into the EAV/CR Model**

As described previously, we have currently implemented five HBP databases—ORDB, OdorDB, CellPropDB, NeuronDB, and ModelDB—in different stages of development and deployment. The first four of these databases were originally written using a conventional relational database design. As it became clear that the underlying EAV data model could be adapted (in EAV/CR form) to handle neuroscience data, we decided to port all four databases to this design. As a result, the latest versions of all five databases are now incorporated into a single EAV/CR database system.31 From a technical standpoint, all the data are pooled. From the user standpoint, each...
database appears to be a separate entity with the same look, feel, and capabilities that it had before the port to EAV/CR. We feel that the successful porting of these very different databases to a single EAV/CR structure represents a good pilot demonstration of the flexibility of the EAV/CR approach.

**Discussion**

This section discusses certain issues that arose in the course of our HBP work.

**SenseLab as One Pilot Model of the Human Brain Project Vision**

The goal of the national HBP is to develop enabling technology for neuroscience research. A central component will involve developing data repositories for well-defined sets of experimental results at many levels of neuroscience, as well as approaches that allow those data to be analyzed in an increasingly integrated fashion. One important way to accomplish this integration will involve building computer-based models that help explain experimental phenomena and make predictions that can be tested in the laboratory. In addition, to make the whole approach as robust as possible, basic informatics research will need to play an important role.

SenseLab provides one model of how these goals can be approached. To help illustrate this pilot model of the HBP vision, Figure 11 shows schematically how SenseLab may evolve:

- We are currently building databases and tools at two major levels of olfactory research—the olfactory receptor (OR) level and the neuron level.
- We are already building and using computer-based models at each of these levels.
- At the OR level, Figure 11 shows a number of ways in which the work may evolve. These include developing a database of OR models that can be tested against different ligands (odor molecules). We have one OR model now. Included in these databases could be the results of running each OR model with various ligands, as well as the results obtained when each OR is exposed to each ligand in the laboratory. In addition, as Figure 11 indicates, there are a variety of other types of data that could be included in this cluster of OR-related databases.

**Neuronal Databases:**

- **CellPropDB** properties of neuronal cell types
- **CircuitDB** models of neuronal circuits (future)
- **ModelDB** models of neurons and neuronal compartments

**Olfactory Receptor (OR) Databases:**

- **ORDB** olfactory receptors
- **OdorDB** odor molecules
- **ORmodelDB** molecular models of ORs (future)

Possible extensions to the OR database cluster:
1. mapping ORs to areas of the olfactory bulb
2. olfactory binding proteins
3. second messenger pathways
4. phylogenetic trees

**Figure 11** An outline of the current SenseLab project, with potential future extensions.
At the neuron level, we plan to extend our current work to include models of neuronal circuits.

In addition, an important component of SenseLab as an HBP model is the underlying informatics research that complements and strengthens the work to develop these databases and tools.

Clinical and Other Real-world Applications of SenseLab

Although the current focus of SenseLab’s activities is on basic neuroscience and neuroinformatics research, there are a range of potential clinical and other real-world applications of this work. For example, the biological olfactory system is a much more powerful and sensitive chemical sensor than any currently manufactured device. As the olfactory system’s mechanisms of operation become better understood at a molecular level, there will be the potential to incorporate this knowledge into much more powerful artificial sensors and neural networks.

From a clinical standpoint, a fully populated database built on NeuronDB’s current design has the potential to be a valuable resource for the pharmaceutical industry as it develops therapeutic drugs that act on the central nervous system. The foundation for such drug development involves knowing what transmitters, receptors, and channels operate across many different nerve cells. NeuronDB is currently designed to collect this very information and support the flexible searching and querying of it.

Building Tools that Support the Investigator vs. Building Pilot Resources that Support the Field

At the beginning of our HBP work, roughly six years ago, we explored the development of informatics tools specifically designed to support the experimental capture, storage, and analysis of experimental data in the laboratory. In this preliminary work, we collaborated with individual researchers to help build these tools.

This process turned out to be very time-intensive. It required that the informatics developer become very familiar with the research project of the investigator. In addition, our evolving recognition of the researcher’s needs provided many opportunities to refine and customize the approach. It was hard to draw the line as to how much support and customization was enough. In addition, it became apparent that the nature of the experiments would typically evolve over time, requiring a corresponding evolution of the various informatics tools that supported that research. Also, since each researcher performed different projects, the needs of each researcher were different.

At the same time, a variety of commercially available tools could be used to meet many of the researchers’ basic informatics needs. Developing customized tools, even for a small handful of investigators, was a major time commitment and a potentially frustrating task, since the researcher could always think of additional capabilities that might be useful.

As a result of factors such as these, we evolved our present approach of supporting olfactory research by developing resources that could serve the field as a whole. The focus is therefore on meeting general needs that could best be met by developing centralized capabilities. We refine these tools by using them to support the research efforts of our laboratories, but this is very different from developing customized tools to support the specific informatics needs of individual research projects.

References