ACADEMIC PRESS

Journal of Structural Biology 138 (2002) 145–155

Journal of Structural Biology

www.academicpress.com

A cell-centered database for electron tomographic data

Maryann E. Martone,[a,b,*] Amarnath Gupta,[b] Mona Wong,[a] Xufei Qian,[b] Gina Sosinsky,[a,b] Bertram Ludäscher,[b] and Mark H. Ellisman[a,b]

[a] *National Center for Microscopy and Imaging Research, Center for Research in Biological Structure and Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093-0608, USA*

[b] *San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92093-0505, USA*

Received 27 December 2001; and in revised form 4 March 2002

Abstract

Electron tomography is providing a wealth of 3D structural data on biological components ranging from molecules to cells. We are developing a web-accessible database tailored to high-resolution cellular level structural and protein localization data derived from electron tomography. The Cell Centered Database or CCDB is built on an object-relational framework using Oracle 8i and is housed on a server at the San Diego Supercomputer Center at the University of California, San Diego. Data can be deposited and accessed via a web interface. Each volume reconstruction is stored with a full set of descriptors along with tilt images and any derived products such as segmented objects and animations. Tomographic data are supplemented by high-resolution light microscopic data in order to provide correlated data on higher-order cellular and tissue structure. Every object segmented from a reconstruction is included as a distinct entity in the database along with measurements such as volume, surface area, diameter, and length and amount of protein labeling, allowing the querying of image-specific attributes. Data sets obtained in response to a CCDB query are retrieved via the Storage Resource Broker, a data management system for transparent access to local and distributed data collections. The CCDB is designed to provide a resource for structural biologists and to make tomographic data sets available to the scientific community at large.

Keywords: Electron tomography; Neuroinformatics; High-voltage electron microscopy; 3D reconstruction

1. Introduction

Electron tomography provides a rich source of structural data, filling in imaging gaps between light and electron microscopy, on the one hand (Martone et al., 2000; Wilson et al., 1992), and molecular and cellular structure, on the other (McEwen and Frank, 2001). When combined with high-voltage electron microscopy or serial section approaches, electron tomography can be used to derive 3D reconstructions of relatively large expanses of cellular and subcellular processes (Ladinsky et al., 1994, 1999; Marsh et al., 2001; Martone et al., 2000; Soto et al., 1994; Wilson et al., 1992). The breadth of these reconstructions is approaching that of light microscopy, while still revealing the exquisite, high-resolution detail typically found in the best electron micrographs of ultrathin sections. Electron tomography can also be used to resolve the substructure of individual macromolecules and so provides a natural bridge between molecular and cellular imaging methods (Harlow et al., 1998, 2001; McEwen and Frank, 2001; Taylor et al., 1997, 1999). Researchers are employing tomography not only to investigate the structure of purified protein complexes, but also to investigate their structure in situ, fitting the lower resolution tomographic data with atomic models or molecular envelopes obtained by molecular microscopy and X-ray crystallography (McEwen and Frank, 2001; Taylor et al., 1999). Others are using the superior resolution of cell-level tomograms to search for molecular signatures of proteins in complex cellular environments based on an understanding of protein structure, without the need for protein-specific stains (Bohm et al., 2000; Koster et al., 1997).

Even within the range of structures traditionally studied by electron microscopy, tomography has proven

* Corresponding author. Fax: +619-534-7497.
E-mail address: mmartone@ucsd.edu (M.E. Martone).

1047-8477/02/$ - see front matter
PII: S1047-8477(02)00006-0

an invaluable tool. The superior resolution of computed tomographic slices in the axial dimension compared to physical sectioning has enhanced our understanding of structures on the order of 5–40 nm, a range of structures typically obscured by the thickness of physical sections (Lenzi et al., 1999, 2001). This attribute, combined with the ability to derive 3D data on cellular and subcellular structure, has led to new insights even into such well-studied structures as mitochondria (Mannella, 2001; Mannella et al., 1997; Perkins et al., 1997a).

The data sets created by the tomography community represent an important and unique source of structural information to biologists. To increase access to tomographic data, we are creating a database specifically for the purpose of making electron tomography data sets of cellular structures available to the scientific community. The Cell Centered Database or CCDB (http://ncmir.ucsd.edu/CCDB/) contains both morphological and protein localization data, derived principally from electron tomography. Tomographic data are supplemented by high-resolution light microscopic data derived from laser-scanning confocal and multiphoton microscopy, in order to provide correlated data on higher-order cellular and tissue structure. Resources such as the CCDB will increase the availability of tomographic data sets and provide a repository of useful data and information for those interested in tomographic investigations.

2. Database design

2.1. CCDB schema

The CCDB (http://ncmir.ucsd.edu/CCDB/) was designed for around the types of data produced at the National Center for Microscopy and Imaging Research (NCMIR) (http://www.ncmir.ucsd.edu). Electron tomography is performed on cellular structures contained within thick sections, 0.5–4 μm (Capani et al., 2001; Lenzi et al., 1999; Martone et al., 1999, 2000; Shoop et al., 1999). Electron tomographic data are produced from single or double axis tilt series taken primarily with an intermediate voltage electron microscope (JEOL 4000EX). The CCDB also contains reconstructions from thicker sections (2–5 μm) imaged on either the 1.25-MeV Hitachi high-voltage electron microscope at the National Physiological Institutes in Okazaki, Japan, or the 3-MeV ultrahigh-voltage electron microscope at Osaka University, Japan. Data from the Osaka UHVEM is obtained either during visits to Osaka University or by remote usage of the UHVEM using a specially designed telemicroscopy kiosk currently housed at NCMIR (Takaoka et al., 2000). The image processing, 3D tomographic reconstruction methodology and tools used for electron tomography at NCMIR are described in Perkins et al. (1997b). Double-tilt axis tomograms are processed using the IMOD package developed by the Boulder Laboratory for Three-Dimensional Fine Structure at the University of Colorado (Kremer et al., 1996).

The database is being developed using Oracle 8i and is housed on a 16 node Sun E10K server at the San Diego Supercomputer Center at the University of California, San Diego. The current schema contains over 80 tables which include experimental, imaging, and reconstruction details as well as the results of any analysis of morphological and protein localization data. An overview of the schema organization is shown in Fig. 1 and an Entity Relationship (ER) diagram (ER/Studio, Embarcadero Technologies, San Francisco, CA, USA) showing the relationships between CCDB tables can be viewed at http://ncmir.ucsd.edu/CCDB/diagram.html. All steps of 3D reconstruction are modeled in the database, from specimen preparation to the final analysis. The images are stored separately from the descriptive data, using the Storage Resource Broker for storage and retrieval of the image data sets (described below). Oracle was chosen as a platform because its Object Relational modeling capability allows us to program any analysis

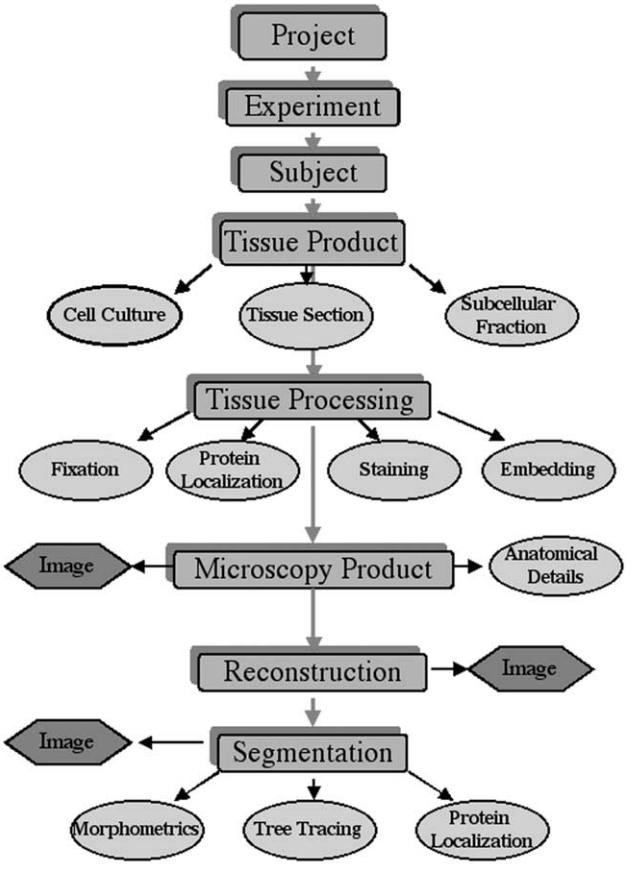

Fig. 1. Schematic view of organization of the CCDB showing a summary of the relationships between the major tables. Image data are included at the level of Microscopy Product (see text for description), Reconstruction, and Segmentation. All images have description and analysis fields attached.

or data comparison methods into the database itself. This functionality allows us to move beyond simple retrieval queries to allow us, for example, to create a "compare two protein distributions" function that runs as an integral method from within the database. The CCDB is currently under development but is expected to be available on the web by June 2002.

Examples of the types of tomographic data sets that are included in the CCDB are shown in Fig. 2. These include relatively low-resolution tomograms of large structures such as selectively stained neuronal spiny dendrites (Figs. 2A–C), as well as higher-resolution tomograms of multicomponent structures like the Node of Ranvier (Figs. 2D and E). The CCDB contains protein localization data derived from immunocytochemistry, enzyme histochemistry, protein-specific dyes, drugs and toxins, and genetic-tagging techniques. An example of correlated light and electron tomographic localization of the gap junction protein, connexin 43, based on a genetically introduced marker is shown in Figs. 2F and G.

The CCDB is not limited to tomographic data but also includes data sets obtained from light microscopy, particularly confocal and multiphoton data, stored as either single 2D images, stacks of optical sections or time series data, when live imaging is performed. Such data are included because many of the tomographic reconstructions are correlated with light microscopic images to provide a larger context for the more detailed electron microscopic reconstructions. Correlated light and electron microscopic data sets are cross-referenced and a map detailing the relationship between the two volumes is included.

The CCDB was designed for cell-level information from tissue, cultured cells, and subcellular fractions, regardless of the type of tissue being studied. In addition, because a significant portion of the tomographic work performed at NCMIR concerns neuronal structures such as spiny dendrites and synaptic structures, the CCDB contains several tables specific for neuronal data. For example, morphological information on the structure of individual nerve cells is included from cells filled with fluorescent intracellular dyes (see Bushong et al., 2002, for details). Filled neurons are stored as a series of optical slices, both with and without deconvolution, and also as branching tree structures traced using Neurolucida (Microbrightfield, Colchester, VT, USA).

2.2. Data modeling

A major part of the effort involved in developing a database is spent in modeling the type of information that is to be included in the database and how it is to be queried (Carazo and Stelzer, 1999; Lindek et al., 1999). In order to accommodate the varied light and electron microscopic reconstruction types, the concept of the "microscopy product" was created (Fig. 3). The microscopy product refers to the set of images taken from the microscope, either as a series of 2D images, e.g., a tilt series, or as a single file containing multiple images, e.g., a confocal file containing optical sections. These images are stored individually but linked together into a set which is used for the reconstruction. Each microscopy image is stored with key experimental details on specimen preparation and pointers to files containing complete protocols. The computed volumes are stored with the acquisition and processing details required to interpret the results of a reconstruction or to recompute the tomographic volume from the original tilt series.

In order to be able to retrieve data based on the features contained within the data, the CCDB stores the results of any analyses performed on a data set. For example, each object segmented from an image or volume is stored as a separate object along with derived measurements such as surface area, volume, length, and number (see Fig. 4). The "number" attribute refers to a count of entities in a population tagged as a single object, e.g., synaptic vesicles. By representing the segmented objects in this fashion, users can retrieve data sets based on the specific morphometric characteristics of features contained in the reconstruction. For example, the user may query for spiny dendrites with spines longer than 2 μm or for synapses with at least 50 synaptic vesicles.

The CCDB contains images showing the distribution of proteins at the light and electron microscopic levels, obtained using immunocytochemistry or some other labeling technique (Figs. 2F and G). Storing protein localization information in a database is a difficult problem because most methods are not quantitative and considerable variation exists in the intensity of staining results, even within a given experiment. We would like to be able to address queries such as "Find cells expressing high levels of protein X and low levels of protein Y." Toward that goal, in the current version of the CCDB, users can provide a labeling intensity value or a rank (high, medium, low, or absent) to represent the levels of protein labeling present in a given structure. Neither of these methods is completely satisfactory because they are subjective and relative measurements, but may provide us with at least rudimentary capabilities to query patterns of protein labeling.

A constant concern in creating and maintaining databases of experimental information is the quality of the data stored in the database. At this time, the CCDB will accept both published and unpublished data, and evaluation of the quality and accuracy of morphometric or protein distribution modeling will be up to the user. The data model employed by the CCDB should aid in this process. First, the morphometric data stored with the objects allows the user to compare the statistics of a given data set to other stored and published data to determine whether they fall within expected ranges.

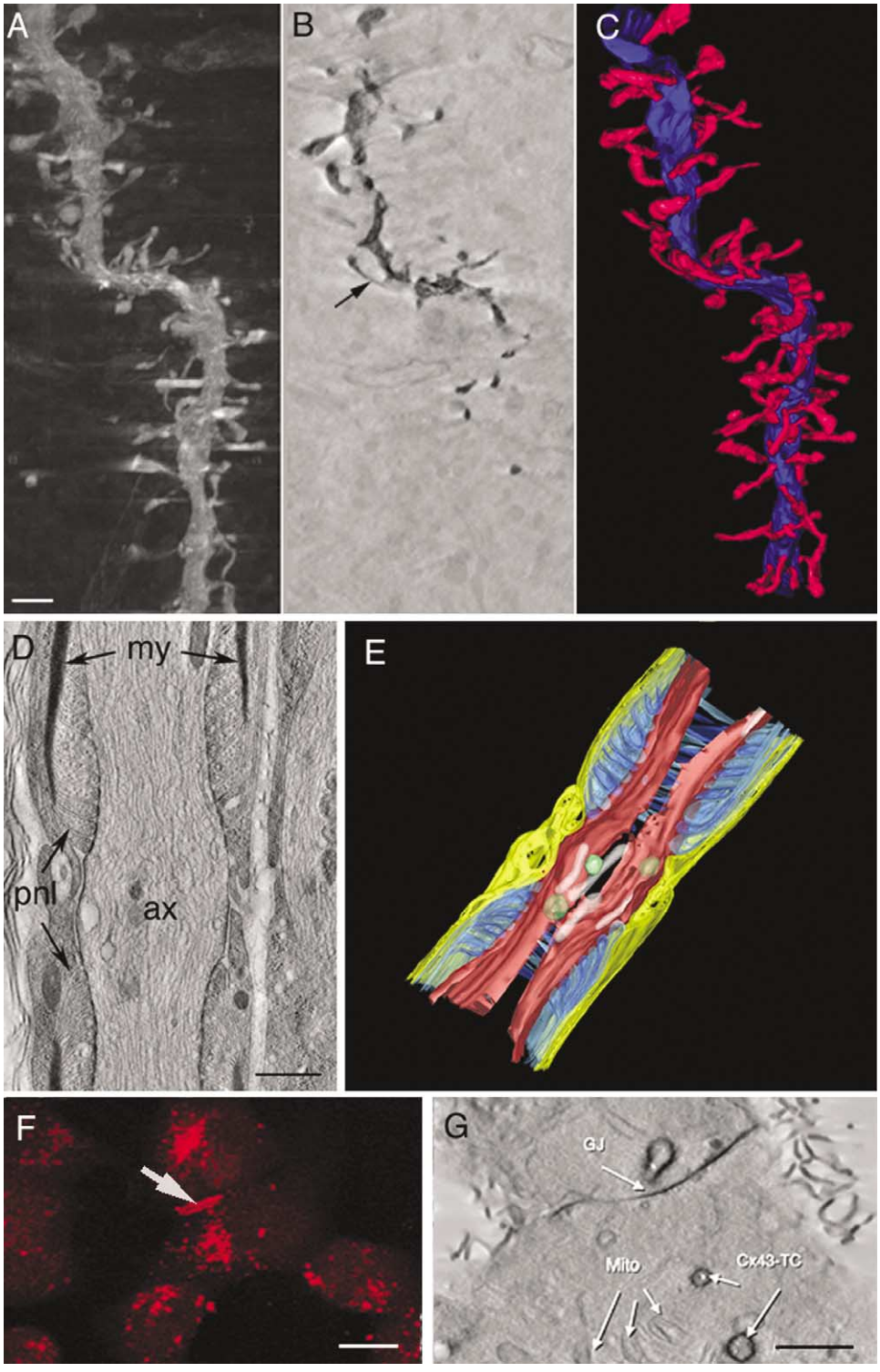

Fig. 2. Examples of data sets contained in the CCDB. (A–C) Tomographic reconstruction of neuronal spiny dendrite from the rat neostriatum stained by intracellular injection of Lucifer yellow followed by photooxidation. This volume was derived from a 4-μm-thick section imaged on the Osaka 3 MeV UHVEM. (A) A maximum intensity projection through the volume. (B) Single computed section through the volume. (C) A surface reconstruction of the segmented volume showing the dendritic shaft in blue and the dendritic spines in red. (D and E) Serial tomographic reconstruction of the Node of Ranvier from three 0.5-μm sections of peripheral nerve. (D) Single computed slice through the node. Many of the components of paranodal glial-axonal junction structures can be resolved, including the axon (ax), the paranodal loops (pnl) and compact myelin (my), indicating a resolution of somewhat better than 100 Å, even at the relatively low magnification required to reconstruct an entire node by serial section electron tomography. (E) Segmented reconstruction of the node showing the major components in different colors. Yellow, compact myelin; red, axolemma; blue, paranodal loops; white, mitochondria; green, intraaxonal vesicles. (F and G) Correlated light and electron microscopic imaging of

Second, the CCDB contains the raw data along with all the imaging and processing steps to allow the accuracy and quality of the final reconstruction to be assessed by an experienced user. Third, the CCDB notes whether or not the data come from published studies. Fourth, the CCDB contains several evaluation tables to allow users users to store estimations of the quality of experimental, imaging, protein labeling, and reconstruction results. Finally, because the interpretation of data is often subjective, users will be able to supply additional or alternative interpretations of a given data set, indexed under their name. In this way, the CCDB can serve as an interactive forum for data interpretation.

2.3. Querying the database

The CCDB can be queried through a web interface. Currently only standard, predefined queries can be processed but the query interface will be updated to support more advanced, user-defined queries. A sample query page is shown in Fig. 5. The database is searchable according to a variety of criteria including the type or size of cellular structure, reconstruction parameters, anatomical region, cell type, or protein. Users may obtain a listing of existing database entries for a given field prior to launching a query.

The database contains pointers to the data files which are stored and accessed via the Storage Resource Broker (SRB: http://www.npaci.edu/DICE/SRB/). The SRB is sophisticated client-server middleware that provides a uniform interface for connecting to data resources over a network. Unlike conventional access methods, e.g., file servers, ftp, or http, SRB is grid-based software providing transparent access to data, relieving the user (in our case the CCDB) from dealing with aspects such as physical location of imaging data, concrete storage devices, and device-dependent access protocols. Thus, regardless of where the data lives, whether in a single location or distributed across several databases, file systems, and high-performance storage systems, SRB provides access to the data via a logical SRB identifier. If multiple copies of the same data are created as a precaution or for efficiency of access, SRB keeps track of the replicas. SRB accomplishes this by creating logical collections of physically distributed data objects (and their copies) that are managed by a central Metadata Catalog (MCAT). MCAT also manages authentication and access control for the data. The CCDB acts as a single client of the SRB/MCAT system, so that separate authentications and accounts do not have to be obtained for each user.

2.4. Displaying the results of queries

Results of queries can be visualized in several ways. The names of volume files are returned along with a thumbnail sketch, a low-resolution QuickTime or AVI movie, and a list of any related files, e.g., tilt images, fiducial mark files (Fig. 5). The user may choose to riffle through the volume slices via the QuickTime movie or to retrieve the actual data set via the SRB. At this time, the CCDB is storing data in the formats used by NCMIR programs (principally SUPRIM, SYNU, and ANALYZE formats). All data in the CCDB are viewable using tools available for download on the CCDB website. Conversion programs to other formats such as MRC, TIFF, and PGM are available either from the NCMIR facility or using the program EM2EM available from Image Science Software GmbH (URL http://www.ImageScience.de). Tomographic data can be viewed as either volume representations or as individual slices using Xvoxtrace, a tool for manual segmentation and visualization of tomographic data developed by Stephan Lamont at NCMIR (Perkins et al., 1997a). Xvoxtrace currently runs on Sun ULTRA (Solaris 2.x), Red Hat Linux 7.x, FreeBSD 4.x, and Silicon Graphics IRIS (IRIX 6.x) and is being ported to Windows systems supporting an X window system server. Xvoxtrace outputs segmented objects as either stacks of contours, volumes, or surfaces. Surface data generated by Xvoxtrace is exported in a format used by the surface visualization program Synu (Hessler et al., 1992). The Synu format can also be converted into other formats such as those supported by Open Inventor. The original Synu program ran only on Silicon Graphics machines, but has recently been rewritten in OpenGL in order to make it more platform independent. Both Synu and Xvoxtrace are freely available and fully documented. Java-based implementations of several of the visualization tools created at NCMIR are currently under development.

2.5. Access to the CCDB

The CCDB is free to all registered users. Access to data stored in the CCDB is provided at three levels: (1) private, (2) local use (i.e., for those with accounts at

genetically engineered gap junction proteins (connexin 43), labeled using ReAsH, an arsenical resorufin derivative (see Gaietta et al., 2002, for details). (F) Confocal light microscopic image of a labeled gap junction between two adjacent cells (arrow). Labeling is also visible in the Golgi apparatus and numerous vesicles. (G) A single computed slice through a tomographic reconstruction of the same two cells is shown in F and G after photooxidation of the ReAsH and processing for electron microscopy. The gap junction (GJ) is clearly labeled, as are transport vesicles or lysosomes (marked as arrows and Cx43-TC) in the cytoplasm. Mitochondria (Mito) and other organelles are clearly visible. Scale bars in A and F = 1 μm; Scale bar in D = 500 nm; Scale bar in G = 5 μm.

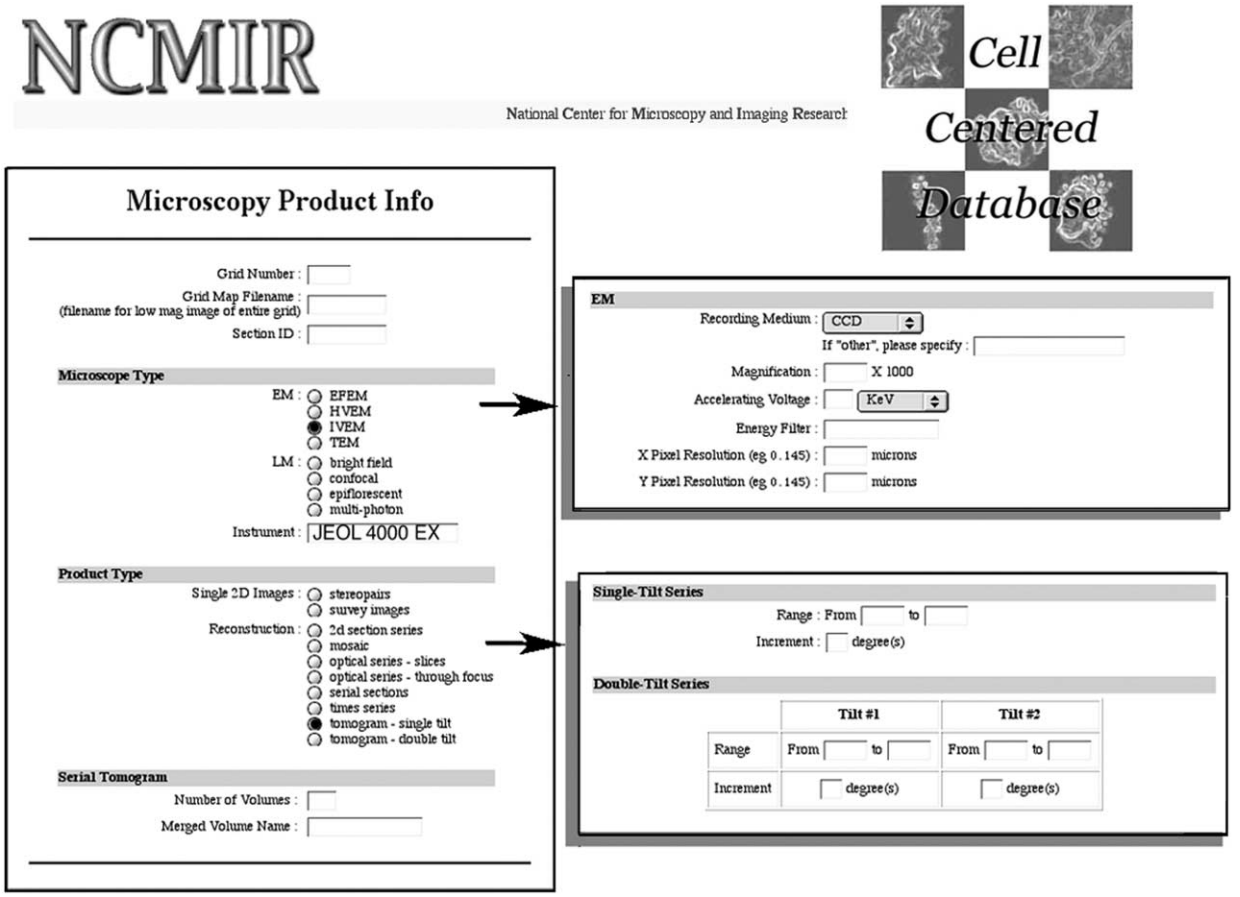

Fig. 3. Data entry forms for some of the imaging parameters contained in the CCDB. To minimize the number of forms seen by the user, the input forms are customized according to the particular conditions. In the "Microscopy Product" form (left-hand menu), the user selects the microscopy and product type, and only the tables appropriate for these selections are presented to the user. The electron microscopy table shown in the upper right represents a version streamlined for manual entry of microscope parameters. A more extensive set of microscopy parameters is read directly from digital images acquired using a CCD camera attached to the microscope. A form with parameters specific to tomographic data acquisition is shown in the lower right. EFEM, energy-filtered transmission electron microscopy; HVEM, high-voltage electron microscopy; IVEM, intermediate voltage electron microscopy; TEM, transmission electron microscopy.

NCMIR), and (3) public use. If data are tagged private, it is accessible only by the author. For local and public access levels, owners of the data may make the data freely available to all with the required permissions, or they may choose to give access only to the descriptive data and low-resolution image views. If a user retrieves a data set tagged in this fashion, an e-mail is sent notifying the owner who then has the choice of granting permission for downloading of the original data, similar to systems employed by other databases (Miller et al., 2001; Roland et al., 2001). Data may only be used for noncommercial purposes and the CCDB and the owner of the data should be acknowledged in any resulting articles.

Submissions from users outside of NCMIR are welcome, including both original data and additional analyses of data sets existing in the CCDB. As described above, the data model is expanded in some areas to accommodate neuronal-specific data, e.g., neuronal tree structures, but the CCDB was designed for cell-level data of any kind. Instructions for depositing data into the CCDB by outside users will be provided on the web site. The amount of data that will be stored with a given reconstruction will vary depending upon the processing stream, but currently, users will be required to store the raw, aligned data used to compute the final volume along with the resulting reconstruction. Such a policy will allow users of the CCDB to perform a reconstruction de novo from existing data according to their needs and provide developers of reconstruction and analytic algorithms access to the primary data. Although only a subset of the description fields will be required, e.g., species, fixation technique, tilt range, and increment, the annotation and storage capabilities of the CCDB will be of benefit to the owner of the data who will have a complete digital and searchable record of the attributes of the particular data set. Thus, the CCDB can serve as an electronic laboratory record manager.

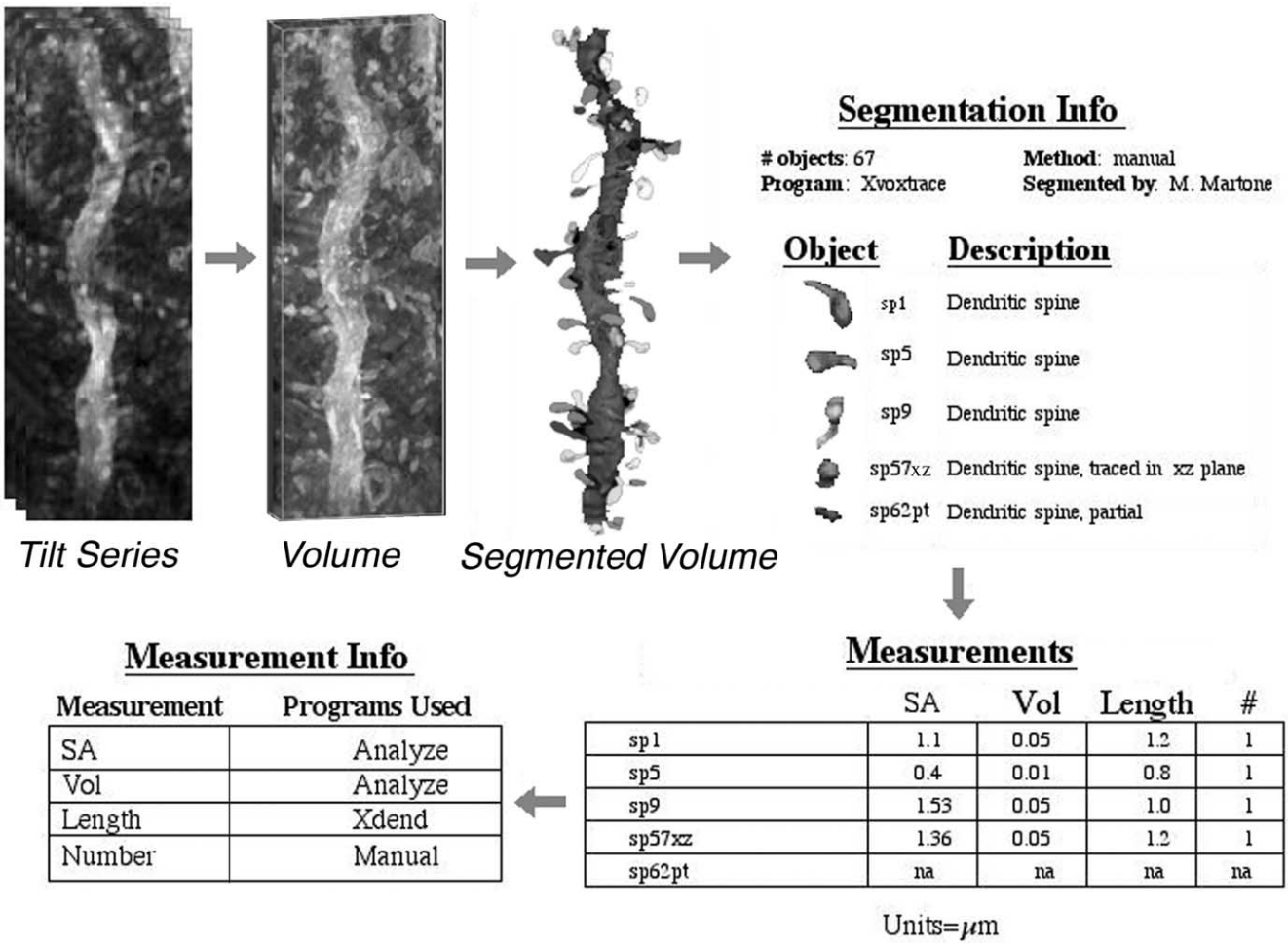

Fig. 4. Data modeling in the CCDB. The data processing stream is modeled from image acquisition through analysis. The images show a spiny dendrite from a rat Purkinje neuron, stained by intracellular injection with Lucifer yellow followed by photooxidation. The displayed images include the tilt series used to create the reconstruction, a maximum intensity projection of the tomographic volume (Volume) and a surface reconstruction derived from the volume reconstruction (Segmented Volume). In this particular segmentation, the dendritic spines, the small protrusions from the dendrite, were segmented individually from the main dendritic shaft. The segmented objects may be listed individually along with a description, associated measurements, and the programs used to generate the measurements. Only a subset of the 67 objects segmented from this particular volume is displayed in this example.

2.6. Extensibility

As more types of tomographic structures are analyzed by CCDB users, there will be a need to extend CCDB beyond its current set of objects and tables. For example, the CCDB does not model spatial relationships, e.g., positions of vesicles within a synaptic cleft captured during synaptic transmission or the position of spines along a dendritic shaft. If we want to extend the current CCDB schema to include such a distribution model, we will take advantage of Oracle's Object Relational modeling capabilities, and create a new set of subclasses of the appropriate existing object class(es) without affecting the rest of the database. To obtain feedback from the scientific community, we will encourage users to send us their suggestions and needs via web forms included on the CCDB website.

3. Utility of the CCDB

The CCDB is just one effort among many to create web-accessible data resources for all types of imaging, from whole organs to molecules. The CCDB complements other cell-level imaging databases such as the BioImage database (http://www.bioimage.org) and related efforts by the ORIAL project of the European Union (http://www.oriel.org) (Carazo and Stelzer, 1999; Carazo et al., 1999; Lindek et al., 1999). The creation of these image databases presents considerable challenges compared to those for sequence and protein structure information because of the heterogeneity and complexity of the primary data. Nevertheless, the effort to create these databases for tomographic data is justified by several immediate and potential benefits. Our discussion below is focused on these benefits and not on some of

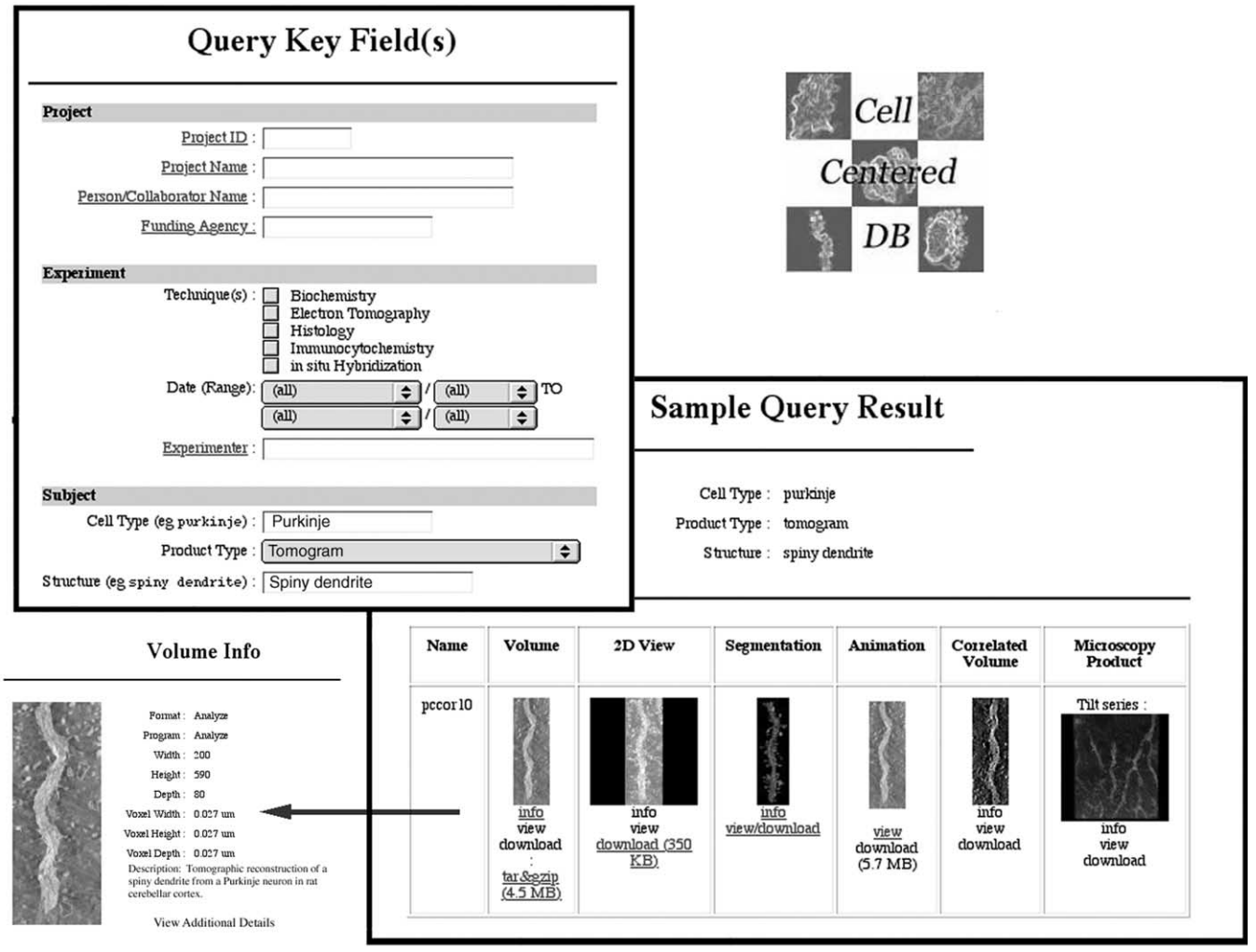

Fig. 5. A sample query and results. When a file is returned from a query, all related files are shown in a table, e.g., correlated volumes, segmented volumes, 2D image views, tilt images, and additional files such as segmentation, measurements, and protocols. The user has the option of viewing text information stored with the file (info), viewing a larger version of the thumbnail sketch or a QT movie (view) or downloading the data set (download).

the well-recognized difficulties involved in creating and maintaining databases and in sharing primary data. Such concerns are significant but have been discussed in detail in several recent reviews of neural imaging databases (Chicurel, 2000; Kotter, 2001; OHBM Commentary, 2001; Roland et al., 2001).

3.1. Providing links to 3D data

Databases like the CCDB can supplement and augment 3D data reported in journal articles. Traditional journal articles are not the ideal way to represent 3D data (Pittet et al., 1999). Both the space limitations on the number of examples that may be displayed and the difficulty in representing 3D data in a 2D format are drawbacks in reporting on tomographic data in the literature. Databases provide a convenient means for storing data referenced in the literature as well as providing access to related data sets and ancillary information.

Databases can also serve as interactive forums for analysis and validation of the data included in them. For example, after retrieving a given data set, a user may deposit additional analyses in the database indexed to the original data. Users can specify whether their analysis agrees with the one on record or provide alternative views. Such alternative views are common when it comes to segmenting anatomical or other structural data, where independent interpretation of boundaries or domains in the 3D structure may not agree. Such inconsistencies have traditionally been viewed as a major impediment to representing imaging data in databases (Kotter, 2001). However, electronic archives of primary and derived data provide the ideal means for uncovering areas where scientific disagreement exists and for visualizing these points of contention directly.

3.2. Richness of tomographic data

Tomographic reconstructions of cellular structures provide unique, high-resolution 3D views of cellular components. Great progress has been made in data collection (Koster et al., 1997), fiducial tracking (Kremer et al., 1996), and computational speed and parallelization of algorithms (Fernández et al., 2002; Smallen et al., 2000), all of which streamline and accelerate the process of tomographic reconstruction to the point where experimenters can go from data collection to reconstruction in less than a day. Analysis of these reconstructions is still far from routine, however, and currently represents the major bottleneck for tomographic studies. Thus, many of these reconstructions contain far more data than are analyzed by a given researcher.

Databases such as the CCDB can make maximal use of rich data like tomographic reconstructions by making the data sets available for additional analyses and by storing the fruits of Herculean efforts such as the extensive reconstructions of the Golgi apparatus (Ladinsky et al., 1999; Marsh et al., 2001) and the Node of Ranvier (Figs. 2D and E). A repository of well-studied structures, such as synapses, dendritic spines, kinetochores, and microtubules, offers a gold mine of structural information to the biological community that can potentially be reanalyzed for different purposes. As more examples of structures such as synapses are added to the repository, additional analyses of structural variability become possible and the relationship of this variability to processes such as synaptic functioning can be modeled. Tomography researchers interested in merging cellular and molecular data will also benefit from having a repository of cell-level structures to search for molecular signatures (Bohm et al., 2000; Koster et al., 1997). In this case, the CCDB can work in concert with molecular databases, e.g., the IIMS database (http://msd.ebi.ac.uk/iims.html) or the PDB, which can provide templates for pattern searches in cellular images.

3.3. Resource for modelers

The development of sophisticated modeling and simulation programs for cellular functions creates a demand for high-resolution and detailed structural data. Modeling programs like MCell (Stiles et al., 2001), Neuron (Hines and Carnevale, 2001), and Genesis (Bower and Beeman, 1998) can accommodate realistic branching structures and surface morphologies such as those contained in the CCDB. Surface reconstructions derived from electron tomographic 3D volumes are now being used as the basis for functional simulations of molecular diffusion and interactions using MCell, a powerful Monte Carlo-based modeling package created for simulation of cellular microdomains (Stiles et al., 2001). Xvoxtrace outputs segmented tomographic data in a form suitable for MCell, which requires the generation of a fine-grained polygon mesh onto which are placed individual molecules or macromolecular complexes. Thus, tomographic reconstructions can provide the structural framework for investigations of molecular dynamics at a cellular scale. An analogous situation occurs with the Protein Data Bank (PDB), which serves as a source of input data of atomic structures for modeling sequence homology and ligand binding and for protein motion calculations (Berman et al., 2000).

3.4. Resource for the tomography community

Beyond providing a service to the scientific community at large, tomographic structure databases can serve as a useful resource for the tomography community. Databases like the CCDB provide a varied source of raw data for those interested in developing new algorithms for reconstruction, segmentation, visualization, or image analysis of electron tomographic data. The utility of databases for spurring analysis tools has already been shown (Laguna et al., 1997; Shindyalov and Bourne, 2001).

For those embarking on tomographic studies, databases of tomographic studies can provide guidelines for setting acquisition parameters for reconstructing a given class of structure. For example, a user may search for reconstructions of the structure of interest or, if such reconstructions are not available, for structures of a similar size. Along these lines, the CCDB has been interfaced directly with the Telescience Portal, an end-to-end tomography application being developed at NCMIR (http://ncmir.ucsd.edu/Telescience/), which provides a centralized web interface to tomography applications including remote microscopy, image viewers, analysis tools, and computational resources. The CCDB can be queried directly from the Telescience Portal to provide guidance in the acquisition and processing of tomographic data. Tomographic data acquired and processed using the Telescience Portal are automatically deposited in the CCDB, when released by the owner.

4. Future directions

Bioinformatics seeks as one of its goals to provide the means to integrate biological information across scales and disciplines. To achieve such a lofty goal, databases like the CCDB, which on their own cover only a small portion of the biological spectrum, will have to be linked to other databases. Creating these linkages for biological data presents a major challenge to the informatics community and one that is proceeding on several fronts (e.g., Gardner et al., 2001; Miller et al., 2001; Rachedi et al., 2000). In our own work, we are linking the CCDB to

databases containing structural and functional information at different scales, from whole brains to molecules, using a novel mediator-based architecture which allows for the incorporation of additional knowledge in order to bridge between sources (Gupta et al., 2000; Ludäscher et al., 2001). The user never interacts directly with a given source, but only with the mediator which queries the relevant sources and returns the results as an integrated view. For example, morphological databases such as the CCDB can be linked to molecular databases like the PDB and physiological sources such as Senselab (http://senselab.med.yale.edu/senselab/). Through the mediator, the CCDB is also linked to additional sources of information such as taxonomies and ontologies like the Unified Medical Language System (http://www.nlm.nih.gov/research/umls/), which provide sources of bridge knowledge and aid in query processing.

Imaging databases like the CCDB are still in their infancy, but it is important for the communities they aim to serve to be involved in their creation and promotion. Their utility will increase as scientists populate these databases and begin to see the possibilities enabled by electronic data representation and access (Kotter, 2001). As they become linked with other web resource through technologies such as database federation, we will be able to navigate through many levels of biological complexity and come closer to the goal of understanding biological systems across scales and functionalities.

Acknowledgments

This work was supported by NIH grants from NCRR RR04050, RR RR08605, and the Human Brain Project DC03192, and NSF grants supporting the National Partnership for Advanced Computational Infrastructure NSF-ASC 97-5249 and MCB-9728338. The authors thank Mr. Thomas Deerinck, Mr. Galen Hand, Mr. Casey Buitenhuys, and Dr. Guido Gaietta for providing some of the images included in this report.

References

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242.

Bohm, J., Frangakis, A.S., Hegerl, R., Nickell, S., Typke, D., Baumeister, W., 2000. From the cover: toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms. Proc. Natl. Acad. Sci. USA 97, 14245–14250.

Bower, J.M., Beeman, D., 1998. The Book of Genesis: Exploring Realistic Neural Models with the General Neural Simulation System. Telos, Santa Clara, CA.

Bushong, E., Martone, M.E., Ellisman, M.H., 2002. Protoplasmic astrocytes in ca1 stratum radiatum occupy separate anatomical domains. J. Neurosci. 22.

Capani, F., Martone, M.E., Deerinck, T.J., Ellisman, M.H., 2001. Selective localization of high concentrations of F-actin in subpopulations of dendritic spines in rat central nervous system: a three-dimensional electron microscopic study. J. Comp. Neurol. 435, 156–170.

Carazo, J.M., Stelzer, E.H., 1999. The bioimage database project: organizing multidimensional biological images in an object-relational database. J. Struct. Biol. 125, 97–102.

Carazo, J.M., Stelzer, E.H., Engel, A., Fita, I., Henn, C., Machtynger, J., McNeil, P., Shotton, D.M., Chagoyen, M., de Alarcon, P.A., Fritsch, R., Heymann, J.B., Kalko, S., Pittet, J.J., Rodriguez-Tome, P., Boudier, T., 1999. Organising multi-dimensional biological image information: the bioimage database. Nucleic Acids Res. 27, 280–283.

Chicurel, M., 2000. Databasing the brain. Nature 406, 822–825.

Fernández, J.-J., Lawrence, A.F., Roca, J., García, I., Ellisman, M.H., Carazo, J.-M., 2002. High-performance electron tomography of complex biological specimens. J. Struct. Biol. 138, 6–20.

Gaietta, G., Deerinck, T.J., Adams, S.R., Bouwer, J., Tour, O., Laird, D.W., Sosinsky, G.E., Tsien, R.Y., Ellisman, M.H., 2002. Multicolor and electron microscopic imaging of connexin trafficking. Science 296, 503–507.

Gardner, D., Abato, M., Knuth, K.H., DeBellis, R., Erde, S.M., 2001. Dynamic publication model for neurophysiology databases. Philos. Trans. Roy. Soc. London B 356, 1229–1247.

Gupta, A., Ludäscher, B., Martone, M.E., 2000. Knowledge-based integration of neuroscience data sources. In: Proceedings of the 12th International Conference on Scientific and Statistical Database Management (SSDBM'00). IEEE Computer Society Press, Silver Spring, MD.

Harlow, M., Ress, D., Koster, A., Marshall, R.M., Schwarz, M., McMahan, U.J., 1998. Dissection of active zones at the neuromuscular junction by em tomography. J. Physiol. Paris 92, 75–78.

Harlow, M.L., Ress, D., Stoschek, A., Marshall, R.M., McMahan, U.J., 2001. The architecture of active zone material at the frog's neuromuscular junction. Nature 409, 479–484.

Hessler, D., Young, S.J., Carragher, B.O., Martone, M.E., Lamont, S., Whittaker, M., Milligan, R.A., Masliah, E., Hinshaw, J.E., Ellisman, M.H., 1992. Programs for visualization in three-dimensional microscopy. Neuroimage 1, 55–67.

Hines, M.L., Carnevale, N.T., 2001. Neuron: a tool for neuroscientists. Neuroscientist 7, 123–135.

Koster, A.J., Grimm, R., Typke, D., Hegerl, R., Stoschek, A., Walz, J., Baumeister, W., 1997. Perspectives of molecular and cellular electron tomography. J. Struct. Biol. 120, 276–308.

Kotter, R., 2001. Neuroscience databases: tools for exploring brain structure–function relationships. Philos. Trans. Roy. Soc. London B 356, 1111–1120.

Kremer, J.R., Mastronarde, D.N., McIntosh, J.R., 1996. Computer visualization of three-dimensional image data using imod. J. Struct. Biol. 116, 71–76.

Ladinsky, M.S., Kremer, J.R., Furcinitti, P.S., McIntosh, J.R., Howell, K.E., 1994. Hvem tomography of the trans-golgi network: structural insights and identification of a lace-like vesicle coat. J. Cell Biol. 127, 29–38.

Ladinsky, M.S., Mastronarde, D.N., McIntosh, J.R., Howell, K.E., Staehelin, L.A., 1999. Golgi structure in three dimensions: functional insights from the normal rat kidney cell. J. Cell Biol. 144, 1135–1149.

Laguna, P., Mark, R., Goldberger, A., Moody, G., 1997. A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg. Computers Cardio. 24, 295–298.

Lenzi, D., Runyeon, J.W., Crum, J., Ellisman, M.H., Roberts, W.M., 1999. Synaptic vesicle populations in saccular hair cells reconstructed by electron tomography. J. Neurosci. 19, 119–132.

Lenzi, D., von Gersdorff, H., 2001. Structure suggests function: the case for synaptic ribbons as exocytotic nanomachines. Bioessays 23, 831–840.

Lindek, S., Fritsch, R., Machtynge, J., de Alarcon, P.A., Chagoyen, M., 1999. Design and realization of an on-line database for multidimensional microscopic images of biological specimens. J. Struct. Biol. 125, 103–111.

Ludäscher, B., Gupta, A., Martone, M.E., 2001. Model-based mediation with domain maps. In: Proceedings of the 17th International Conference on Data Engineering, April 2–6, 2001, Heidelberg, Germany. IEEE Computer Society Press, Silver Spring, MD.

Mannella, C.A., 2001. Application of electron tomography to mitochondrial research. Methods Cell Biol. 65, 245–256.

Mannella, C.A., Marko, M., Buttle, K., 1997. Reconsidering mitochondrial structure: new views of an old organelle. Trends Biochem. Sci. 22, 37–38.

Marsh, B.J., Mastronarde, D.N., McIntosh, J.R., Howell, K.E., 2001. Structural evidence for multiple transport mechanisms through the golgi in the pancreatic beta-cell line, hit-t15. Biochem. Soc. Trans. 29, 461–467.

Martone, M.E., Deerinck, T.J., Yamada, N., Bushong, E., Ellisman, M.H., 2000. Correlated 3d light and electron microscopy: use of high voltage electron microscopy and electron tomography for imaging large biological structures. J. Histotechnol. 23, 261–270.

Martone, M.E., Deerinck, T.J., Young, S.J., Ellisman, M.H., 1999. Three dimensional protein localization using high voltage electron microscopy. Acta Histochem. Cytochem. 32, 35–43.

McEwen, B.F., Frank, J., 2001. Electron tomographic and other approaches for imaging molecular machines. Curr. Opin. Neurobiol. 11, 594–600.

Miller, P.L., Nadkarni, P., Singer, M., Marenco, L., Hines, M., Shepherd, G., 2001. Integration of multidisciplinary sensory data: a pilot model of the human brain project approach. J. Am. Med. Inform. Assoc. 8, 34–48.

OHBM Commentary, 2001. Neuroimaging databases. Science 292, 1673–1676.

Perkins, G., Renken, C., Martone, M.E., Young, S.J., Ellisman, M., Frey, T., 1997a. Electron tomography of neuronal mitochondria: three-dimensional structure and organization of cristae and membrane contacts. J. Struct. Biol. 119, 260–272.

Perkins, G.A., Renken, C.W., Song, J.Y., Frey, T.G., Young, S.J., Lamont, S., Martone, M.E., Lindsey, S., Ellisman, M.H., 1997b. Electron tomography of large, multicomponent biological structures. J. Struct. Biol. 120, 219–227.

Pittet, J.J., Henn, C., Engel, A., Heymann, J.B., 1999. Visualizing 3d data obtained from microscopy on the internet. J. Struct. Biol. 125, 123–132.

Rachedi, A., Rebhan, M., Xue, H., 2000. Gabaagent: a system for integrating data on gaba receptors. Bioinformatics 16, 301–312.

Roland, P., Svensson, G., Lindeberg, T., Risch, T., Baumann, P., Dehmel, A., Frederiksson, J., Halldorson, H., Forsberg, L., Young, J., Zilles, K., 2001. A database generator for human brain imaging. Trends Neurosci. 24, 562–564.

Shindyalov, I.N., Bourne, P.E., 2001. A database and tools for 3-d protein structure comparison and alignment using the combinatorial extension (ce) algorithm. Nucleic Acids Res. 29, 228–229.

Shoop, R.D., Martone, M.E., Yamada, N., Ellisman, M.H., Berg, D.K., 1999. Neuronal acetylcholine receptors with alpha7 subunits are concentrated on somatic spines for synaptic signaling in embryonic chick ciliary ganglia. J. Neurosci. 19, 692–704.

Smallen, S., Walfredo, C., Frey, J., Berman, F., Wolski, R., Su, M.-H., Kesselman, C., Young, S., Ellisman, M., 2000. Combining workstations and supercomputers to support grid applications: the parallel tomography experience. In: Proceedings of the 9th Heterogenous Computing Workshop, May 2000.

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

Stiles, J.R., Bartol, T.M., Salpeter, M.M., Salpeter, E.E., Sejnowski, T.J., 2001. Synaptic variability: new insights from reconstructions and monte carlo simulations with mcell. In: Cowan, W.M., Sudhof, T.C., Stevens, C.F. (Eds.), Synapses. Johns Hopkins University Press, Baltimore, pp. 681–731.

Takaoka, A., Yoshida, K., Mori, H., Hayashi, S., Young, S.J., Ellisman, M.H., 2000. International telemicroscopy with a 3 mv ultrahigh voltage electron microscope. Ultramicroscopy 83, 93–101.

Taylor, K.A., Schmitz, H., Reedy, M.C., Goldman, Y.E., Franzini-Armstrong, C., Sasaki, H., Tregear, R.T., Poole, K., Lucaveche, C., Edwards, J., Chen, L.F., Winkler, H., Reedy, M.K., 1999. Tomographic 3d reconstruction of quick-frozen, ca^{2+}-activated contracting insect flight muscle. Cell 99, 421–431.

Taylor, K.A., Tang, J., Cheng, Y., Winkler, H., 1997. The use of electron tomography for structural analysis of disordered protein arrays. J. Struct. Biol. 120, 372–386.

Wilson, C.J., Mastronarde, D.N., McEwen, B., Frank, J., 1992. Measurement of neuronal surface area using high-voltage electron microscope tomography. Neuroimage 1, 11–22.