

LIVE CELL FLUORESCENCE IMAGING AND IMAGE INFORMATICS

Walter A. Carrington
Physiology Department
University of Massachusetts Medical School
222 Maple Avenue
Shrewsbury, MA 01545 USA
email: Walter.Carrington@umassmed.edu

KEY WORDS: Live cell, 3-d, 4-d, fluorescence, high speed, microscopy, grid computing, informatics.

There is an impending deluge of data from 2-d and 3-d fluorescence imaging of live cells. Advances in cameras and probes such as GFP will make it routine to obtain hundreds or thousands of 3-d images or tens of thousands of 2-d images from a single live cell. Combined with larger, faster cameras this will result in tens or hundreds of TeraBytes of data per year that must be stored, processed, visualized and analyzed. This impending deluge of data threatens to overwhelm the software and hardware infrastructure of groups that do live cell imaging.

In our own laboratory, a new high speed wide field 3-d microscope based on a 640x480 pixel, 94 frame per second CCD camera has recently become operational. This system is capable of producing up to 25 gigabytes of 2-d or 3-d data in 7 minutes from a single dual labeled live cell. Other technologies such as spinning disk confocal or multispot multiphoton microscopes are also capable of producing extremely large data sets from live samples. With modest streamlining or automation of the biological experiments data rates of one Terabyte per day are plausible. Biological laboratories that acquire a relatively modestly priced commercial microscope system may find that a supercomputer center is needed to handle the resulting data.

Experimenter's lab notebooks no longer suffice to organize the images and the information about its acquisition, processing and analysis. As data sets become larger, automation and automated error checking becomes more necessary. This requires that information about the data (metadata) must be more detailed. Methods for defining, organizing and maintaining the image metadata (information about the data) must be developed and standardized. All the information required for deriving scientifically valid conclusions from the images must be stored in a database of metadata, including optical configuration (filters used, objective lens, etc.), the biological preparation (cell type, fluorophore, etc.), image processing steps and the results of interactive analysis and visualization. There are projects that provide excellent starting points for this task— the Open Microscopy Environment (OME, at <http://openmicroscopy.org>) and the Biomedical Informatics Research Network (BIRN; <http://ncmir.ucsd.edu>). Unfortunately, there is more than one such project; standardization is necessary.

A framework for streamlining or automating the data analysis workflow must be developed. Grid computing provides a unified way of accessing a variety of computational platforms across administrative and institutional domains. Software must be designed to be adaptive – to adapt to the problem size, characteristics of the computers available and load on the computer system. Grid computing using standard grid middleware provides a way to organize the storage and processing of these volumes of data over computational grids.