# SANDIA REPORT

SAND2008-6522
Unlimited Release
Printed Sept. 2008

# High-Throughput Proteomics: Optical Approaches

George S. Davidson

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico  87185 and Livermore, California  94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

**Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.**

**NOTICE:** This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2008-6522

Unlimited Release

Printed September 2008

# High-Throughput Proteomics: Optical Approaches

George S. Davidson
Computation, Computers, Information and Mathematics

Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1316

## Abstract

Realistic cell models could greatly accelerate our ability to engineer biochemical pathways and the production of valuable organic products, which would be of great use in the development of biofuels, pharmaceuticals, and the crops for the next green revolution. However, this level of engineering will require a great deal more knowledge about the mechanisms of life than is currently available. In particular, we need to understand the interactome (which proteins interact) as it is situated in the three dimensional geometry of the cell (i.e., a situated interactome), and the regulation/dynamics of these interactions. Methods for optical proteomics have become available that allow the monitoring and even disruption/control of interacting proteins in living cells. Here, a range of these methods is reviewed with respect to their role in elucidating the interactome and the relevant spatial localizations. Development of these technologies and their integration into the core competencies of research organizations can position whole institutions and teams of researchers to lead in both the fundamental science and the engineering applications of cellular biology. That leadership could be particularly important with respect to problems of national urgency centered around security, biofuels, and healthcare.

# Contents

Introduction........ 7
How are we to lay the foundation of predictive cellular modeling?........ 8
Optical proteomics........ 10
  Strategies and the science accessible to optical proteomics........ 11
    Protein localizations and manipulations........ 11
      Förster resonance energy transfer (FRET) interactions........ 12
      Phosphorylated, circularly permuted fluorescent proteins (cpFP)........ 12
      Inference by observing network responses to photo-sensitizers........ 13
  Automation for high throughput laboratories........ 17
  Use of the measurements and cell models........ 18
Discussion and recommendations........ 22
Acknowledgements........ 22
References........ 23

# Introduction

A very deep understanding of cellular machinery will be required to successfully model cells as an aid to bioengineering for biofuel, pharmaceutical, or other industrial applications. To develop these models, we will require a great deal of knowledge about which proteins interact, and how they localize throughout the cell. High-throughput methods are beginning to provide the required data, albeit, still at relatively coarse scales; see (Brown and Davidson 2007) for a survey and discussion of implications for modeling. This report covers an LDRD investigation of new approaches and technologies applying optical methods to proteomic questions and research. For example, breakthroughs ranging from new applications of genetically encoded fluorescent proteins to new super-resolution microscopy methods that will eventually allow the measurement of protein interactions at very fine scales (ultimately down to a few nanometers, but currently 20-30 nanometers—well below the diffraction limit). Considerable work remains to be done before these methods can be matured into high-throughput, production tools, but the way forward is becoming clearer.

The role of DNA looms large in our minds when we think about how cells work; so much so, that it is easy to overlook the important role played by the rest of cell, as a whole, *to interpret* the genetic information, and to realize the chemical structures that eventually transform one cell into daughter cells.

Hence, we must *simultaneously* consider the encoding of genes and the implicit mechanical realization of that code, which is an interpreting system. That realization *is* the cellular architecture, the molecules and their spatial arrangements and coordinated interactions. This arrangement of molecules continuously transform the encoded information into the physical structures of the cell, and thereby into copies of itself.

Knowing the collection of expressed proteins, the proteome, is essential, if we are ever to understand and model cells at that level. However, we must know where the proteins reside, and when they move between cellular components. Because they are embedded in signaling, and metabolic chains, we will not really understand the cell as an interpreter until we can model the complete set of these interactions, the interactome (Rual, Venkatesan et al. 2005).

These interactions, taken as a whole, will almost certainly be so tangled that we will require computer models to help us with the complexity. Constructing these computer models will be a vast research effort, and is currently beyond our capability. However, we are very nearly able to begin addressing one important part of that puzzle: determining the locations (and collocations) and interactions among, at least, some of the critical proteins. The role for light microscopy in determining the protein interactome is the central point of the paper. However, just knowing the interactome is insufficient, the dynamics and spatial localizations of the proteins must also be available for accurate computer modeling of cells and biological processes, and here microscopy will also play a central role.

## How are we to lay the foundation of predictive cellular modeling?

Andreas Wagner, the author of *Robustness and Evolvability in Living Systems* (Wagner 2005), has suggested that by the time we begin to understand most of the mechanisms used in cells, the complexity may be so great that we will no longer find it comprehensible.[1] Nevertheless, we may be able to use computer models integrating an encyclopedic knowledge of the minute details to make accurate predictions for our biological experiments.

This then is the problem: *how are we to complete the task of learning enough to be able to accurately model and simulate actual biological experiments?* Clearly, we are still a long way from that goal. Certainly we must know all of the parts to the finest level required. We must, also, know how these parts are maintained and how they are replaced through metabolism and reproduction. Ultimately, we must understand how manipulations of internal and external molecular concentrations and physical structures will change the otherwise well understood, normal situations. A static understanding will not be sufficient; we will need to capture these details in interacting algorithms that can form the basis of our simulations.

The required models and simulations are not yet achievable, in part because we lack the fundamental knowledge about which proteins collocate and interact with each other, but also because we lack the appropriate conceptual frameworks to deal with the complexity offered by whole cell mechanisms. These are not small holes in our knowledge; they reflect the need to undertake a vast program of research aimed at a fuller understanding of cells. Certainly, the ultimate goal will require building on our knowledge of the biochemistry of nucleic acids, proteins, lipids and sugars as well as numerous small molecules. The increasing combinatorial complexity possible with proteins, sugars and lipids may suggest that our laboratory methods are not yet up to the challenge. However, despite the huge repertoire of proteins, and even with their numerous alternate sequences and folds, we are approaching the capability study the proteome as a whole.

Unfortunately, merely extending the current level of protein annotations to the entire set of translated proteins is unlikely to break barriers on its own. We will still lack understanding of how these proteins interact to regulate biological processes and how the interplay of proteins and the morphology of organelles, cells, tissues, and organs ultimately enable life.

The recent advances in experimental techniques discussed below will allow for an unprecedented quantity of protein interaction data to be collected. Importantly, these data will also be collected within the *in vivo* context of the relevant organelles and membranes. An earlier report (Brown and Davidson 2007) has suggested that computational biologists ought to begin thinking about how to use the coming wealth of information to enable vastly more powerful models of cells as systems of systems. While this paper is emphasizing how to collect that data, it is particularly important to remember that the long term objective is to create predictive models, and that these exciting measurement technologies are just a bridge to that future.

---

[1] Personal communication, 2006.

The scale of this work can be appreciated by considering that relatively simple organisms have several thousand genes. For instance, the eukaryote to be fully sequenced (Cherry, Ball et al. 1997), *Saccharomyces cerevisiae*, yeast, has 5749 protein encoding genes (SGD 2008). Consequently, to look at all possible pairwise protein interactions would require on the order of 18 million experiments. Even restricting these studies to proteins known to be associated with a particular organelle would be massive experiments. For example, There are 681 yeast genes containing 'mitochondria' in their annotations (SGD 2008), which would require over 200,000 experiments. If such studies are to be undertaken, automated, high-throughput methods for both the experiment and the analysis will be required.

If these automated experiments are to resolve interactions to a resolution enabling the identification of the interaction's spatial location (for example, within mitochondrial, or the finer details of nuclear transcription factories, or even specific sides of a membrane), one of the best high-throughput methods, tandem mass spectroscopy (Aebersold and Mann 2003), will probably not be useful. Imaging mass spectroscopy (McDonnell and Heeren 2007) can achieve resolution down to several hundred nanometers; to go lower, ion beams are used for desorption (ablation). Secondary ion mass spectrometry (SIMS) can have 50 nm or less resolution, but images have only been recorded for atomic ions and low mass molecular fragments (McDonnell and Heeren 2007).

Fortunately, new methods of super-resolution microscopy enable sub-diffraction limited resolution; see (Hell 2007) for a recent review of these new methods, and (Davidson 2007) for more detail. High throughput implementations of these new microscopes have not yet been implemented. However, there are examples of fully automated laboratories with multiple robots and automated microscopes. For instance, a facility in Germany has described sixteen such bays that have processed nearly 100 different serial immunofluorescent stains on the same sample (Schubert, Bonnekoh et al. 2006). Scaling up to a facility able to process tens of thousands of samples per day with many hundred to thousand different antibodies will be an engineering challenge, but is likely to be realizable.

Antibody based imaging methods for complete cell protein-protein interactions can be criticized because most of the required antibodies are not available. Further, it might not even be possible to make antibodies against proteins that are highly conserved, or that are expressed at such low levels that antigenic responses can not be induced.

Certainly, phage display methods offer the promise of an alternative to traditional animal-grown antibodies and could well play a critical role in high throughput optical proteomic approaches. In an alternate approach, David Peabody's work with MS2 bacteriophage (Peabody 1997) has led to another possible path toward the production of the required antibodies. MS2 phage can be induced to display antigenic peptides on their surface (van Meerten, Olsthoorn et al. 2001; Brown, Mastico et al. 2002). These coat proteins can be grown with cell-free methods and will self assemble. Further, the dense, repetitive presentation of the polypeptides as part of the coat protein produces extremely strong immune responses, even breaking self immunity (Chackerian 2007). Peabody's methods should allow for the production of antibodies, even for highly conserved proteins. It is expected that either phage-display or MS2 coat proteins can be developed to supply the required antibodies for very large fractions of the required proteins and that these methods, together with the automated optical microscopes, can be used to construct interactomes of sufficient quality and coverage for accurate, computer modeling.

While there are pathways toward large collections of antibodies for immuno-fluorescent staining, the effort and cost associated with acquiring them, and maintaining them, will be considerable. Creating a library of such antibodies would require sacrificing thousands of animals, which would not be popular. Hence, for many applications it may be preferable to have the cells themselves make the required fluorescent tags, which can be achieved, in many cases, by fusing the gene for a natural fluorophore to the native gene encoding the protein of interest. In some cases individual strains are available such that every gene in the organism has been fused with a fluorophore; for example the green fluorescent protein (GFP) tagged strains for *S. cerevisiae* (Howson, Huh et al. 2005).

GFP is not directly suitable for several of the super-resolution methods because, in its most common version, it is not a photo-switched fluorophore. However, alternates such as photo-switched cyan fluorescent protein (Chudakov, Verkhusha et al. 2004; Chudakov, Lukyanov et al. 2007), Kaede (Ando, Hama et al. 2002), and Dronpa (Patterson and Lippincott-Schwartz 2002) are photoswitchable and have been used for super-resolution imaging. These genes can be fused to a protein's gene to be either on the amino or the carboxyl terminus, and often the fusion does not interfere with the target gene's native function.

Hence, various methods are available for marking and visualizing elements of cellular interactomes. How far might we go with these optical methods toward better understanding of proteomics and better modeling?

## Optical proteomics

Here, optical proteomics will mean methods utilizing light microscopes or other optical means to study the interactions and localizations of proteins within cells, cultures, or tissues.

The important issues have to do with 1) strategies applicable to the use of the instruments with respect to what can be known about cellular interactions, 2) how to automate these processes to support very high throughput surveys, 3) the resolution of detection, 4) the use of the experimental measurements to enable new science and better engineering of biological processes, especially with respect to how these measurement can enable better computer modeling of networks and cells. These issues will be examined, below.

## *Strategies and the science accessible to optical proteomics*

A number of strategies and techniques have been developed that use fluorescent signals from particular molecular states within cells. These range from simply detecting the presence of a protein, to tracking cellular dynamics by monitoring the movements of these molecules. Importantly, optical techniques are available to know when tagged molecules are physically close to one another and even ways to monitor the phosphorylation activities of kinases and their target molecules. The following subsections will focus on several novel methods that are being used to reveal, and even control molecular pathways, and which can provide useful data for cellular modeling as well as for traditional bench scale molecular biology.

### Protein localizations and manipulations

Proteins on and within cells can be visually detected using tagged antibodies, for instance with relatively small, fluorescent molecules (for example Cy3 and Cy5, or closely related variants, see (Bates, Huang et al. 2007)) incorporated into antibodies specific for the target protein. This approach can be problematic; certainly the lack of an antibody for the target is a problem, and the tags can be toxic, or there may be difficulties getting the antibody to localize properly. These problems are only compounded for *in vivo* studies. As a result of these concerns, it may be more fruitful to make use of chimeric gene fusions that allow the direct coupling of a gene encoding the protein of interest with a gene for natural or optimized fluorescent proteins. The fusion can be either at the carboxyl, or amino end of the target, and typically includes codons for a few amino acids that act as a spacer between the tagged protein and the fluorescent tag, itself. When this fusion gene is transcribed and translated, the target and the fluorophore will be directly coupled by the spacer, which not only tethers the two parts, but also improves the chance that fusion protein will still fold correctly and be functional for both parts.

In some species, it is possible to get individual strains that each have a different, single gene fused to the gene for the fluorescent tag. For example, a library of 4156 different strains of *S. cerevisiae* has been created so that it is possible to observe optically about 73% of the 5749 protein-coding genes in this important organism (Howson, Huh et al. 2005). If the experimental design does not require a chimeric fusion, perhaps because one wishes to over express the tagged protein, or place it under the control of different promoters, then the fluorescent proteins can carried into the cells with plasmids and expressed independently of nuclear control. In either case, it is possible to include an address tag to target the chimeric protein to a particular organelle; these genomic techniques are very flexible and powerful.

## Förster resonance energy transfer (FRET) interactions

Förster resonance energy transfer (FRET) from one fluorophore to another can occur when the emission spectrum of the donor fluorophore overlaps with the excitation spectrum of the receptor and when the two are physically in contact with each other (Förster 1948). The observation or absence of a FRET signal can, thus, measure when the two fluorophores have come into contact or have separated. This feature can be exploited by attaching FRET pairs to interacting molecules, or to regions near sensor regions in a molecule that undergoes change or refolding, which repositions the FRET pairs. Calcium signaling and second messenger signaling involving cyclic AMP and GMP have been studied with FRET based methods. They are particularly useful for studying conformational changes induced by kinases. In this case the FRET pairs are engineered to be near phosphorylation sites such that a change in phosphorylation and the resulting physical change in molecular shape will move the fluorophores and change the FRET signal.

## Phosphorylated, circularly permuted fluorescent proteins (cpFP)

A particularly intriguing sensor strategy embeds sensor regions within a single fluorescent protein (FP). In this technique called circularly permuted fluorescent proteins the gene for the FP is shuffled at random (or rationally) to combine fragments in a non-native order. Most of these shuffled proteins are no longer fluorescent; however, occasionally the fluorescent property is preserved. While shuffling the genes, it is possible to insert additional base pairs to encode known sensor modules (for example a typical phosphorylation site for kinase being studied). Occasionally, these permuted proteins with insertions will gain or lose fluorescence when the sensor region is activated, which causes a conformational change. Souslova and Chudakov (2007) review a large number of these cpFP sensors, including one developed in their laboratory for $H_20_2$ (Belousov, Fradkov et al. 2006).

> "HyPer (from hydrogen peroxide) consists of the permuted variant of EYFP (cpEYFP) inserted into the regulatory domain of prokaryotic OxyR protein sensitive to hydrogen peroxide. Interaction of HyPer with $H_20_2$ causes oxidation of two cysteine residues of OxyR causing significant conformational changes. This results in conformational changes of the permuted core of the sensor accompanied by changes of its fluorescent properties. […] Using this sensor, it was possible to detect submicromolar oscillations of intracellular hydrogen peroxide in the cytoplasm and mitochondria of cells during apoptosis and growth factor stimulation." (Souslova and Chudakov 2007)

The applications of HyPer should benefit our understanding of the relationships between glucose metabolism, reactive oxygen species (ROS) and apoptosis. Better understanding of these processes will have biomedical and biofuel implications.

## Inference by observing network responses to photo-sensitizers.

Photosensitizers generate reactive oxygen species (ROS) when irradiated, and are particularly useful for inactivating proteins via chromophore-assisted light inactivation (CALI). The ROS can cause sufficient damage to kill the cell, or with a smaller exposure dose, just damage (inactivate) the proteins near the photosensitizer. A genetically encoded photosensitizer, KillerRed (GenBank accession number AY969116), can now be fused with a target protein by constructing a genetic chimera (Bulina, Chadakov et al. 2006).

Killer Red generates both singlet oxygen and superoxides when irradiated with green light (540-580 nm); however, singlet oxygen is the primary damaging agent and cell death is by apoptosis rather than necrosis. Bulina et al. (2006) have demonstrated that with strong illumination (this photosensitizer can kill both prokaryotes (96% of E. coli cells were killed after 10 min of exposure at 1 W/cm$^2$ and almost all after 20 min) and eukaryotes (40-60% of 293-T human kidney cells were killed following 10 min exposure to 5.8 W/cm$^2$).

More interestingly to cell modelers, KillerRed fusions can be used to inactive the protein to which KillerRed protein is fused. This inactivation is via CALI, after an exposure to much less intense light than that required to kill the cell. Bulina et al. (2006), replicating the original demonstration of CALI (Surrey, Elowitz et al. 1998), were able to inactive β–galactosidase fused with KillerRed using a 30 min exposure at 360 m W/cm$^2$. As a second demonstration, they demonstrated CALI with the pleckstrin homology domain of C δ1 (PLCδ1 ) in EGFP-PH fused with KillerRed. A single 10 sec illumination was effective in inactivating the pleckstrin domain without cell death. They estimate that there is at least a 10 to 20-fold difference between the light dose required for CALI and for cell death.

CALI offers the opportunity to transiently or continuously inactivate one or more proteins in a cell, which can be used to synchronize cell growth, or perturb entry into or out of quiescence, or explore pathways and major cell processes like apoptosis. Information from such experiments will be very helpful to cell modelers in constructing and verifying their models, and will generally further deepen our understanding of these molecular processes.

## Protein-fragment complementation assays (PCA)

The hunt for interacting proteins can be rational, as in following and extending the knowledge about a particular network; or it can be totally random as in pull-down experiments using some particular protein; or (only in theory) systematic where all possible interactions are checked. Of course, there are so many possible two and higher, multi-way interactions that the systematic approach is doomed. Rational and random approaches are also flawed, especially with respect to completeness. A new selection based technique (Tarassov, Messier et al. 2008) offers some hope that the combinatorial explosion of possible protein interactions can be winnowed down to large, but achievable systematic studies involving pairs of proteins that must interact for the cell to survive the selection process. While this screening technique is not technically an optical proteomics method, it is discussed here because it offers the opportunity to direct optical methods toward those interactions that are most likely to be important.

The ultimate measure of importance is the direct observation of interacting proteins *in vivo*, for example protein interactomes have been previously constructed using yeast two-hybird screens (Uetz, Giot et al. 2000; Ito, Chiba et al. 2001; Rain, Selig et al. 2001; Giot, Bader et al. 2003; Colland, Jacq et al. 2004; Li, Armstrong et al. 2004; Rual, Venkatesan et al. 2005). However, efforts to find *in vivo* interactions are have hardly provided definitive evidence. Reproducibility rates from yeast two-hybrid have been reported at ~55% for human interactome (Rual, Venkatesan et al. 2005) and false positive rates have been reported to be ~50% (von Mering, Krause et al. 2002). False negative rates have been estimated as high as 85% for yeast two-hybrid and 50% for coaffinity purification with mass spectroscopy (Edwards, Kus et al. 2002; von Mering, Krause et al. 2002).

Clearly, there is an outstanding need for alternative high-throughput approaches and additional data on protein binary interactions (Vidal 2005). The protein-fragment complementation assays (PCA) (Michnick, Ear et al. 2007) offer an *in vivo* alternative, which may give more reproducible results.

In PCA, the genes for complementary parts of a reporter protein are fused to the genes encoding the target proteins. If the target proteins interact, the complementary parts of the report protein are localized and may be able to combine into the active form. Clearly, this approach is not complete as it is not guaranteed that the reporter will be able to become active; however, positive results are strong evidence for the interaction of interest.

Online Mendelian Inheritance in Man entry for DHFR
(McKusick, O'Neill et al. 2008)

Dihydrofolate reductase (EC 1.5.1.3) converts dihydrofolate into tetrahydrofolate, a methyl group shuttle required for the de novo synthesis of purines, thymidylic acid, and certain amino acids. DHFR is inhibited by methotrexate (MTX), a folate analog used as an antineoplastic and immunosuppressive agent.

Tarassov et al. (2008) used complementary parts of a mutant murine dihydrofolate reductase (mDHFR) as the reporter in their PCA-based protein-protein interaction screen in yeast having. This mutant mDHFR is insensitive to methotrexate, which normally lethally inhibits DHFR. In their survival screening method, methotrexate treatment kills all cells except the ones with functioning mDHFR, which by construction requires the combination of the PCA reporter parts; hence demonstrating very close protein localization and probable interaction. Therefore, a surviving colony is evidence for the interaction of the two fusion proteins, which allows whole genome screens for protein interactions to be conducted using this automated colony-based screen.

Terrassov et al. (2008) used this approach to screen all pairs of *S. cerevisiae* genes by creating 4326 haploid strains where each unique strain fused one open reading frame with both N- and C-terminal DHFR fragments. This library of strains covers about 75% of the *S. cerevisiae* genes that encode proteins. They then performed all pairwise crosses to create diploid strains grown on solid phase media. After exposure to methotrexate, colony growth was recorded for each unique protein pair using automated image analysis. They identified 2770 interactions among 1124 expressed proteins, 80% of which had not previously been reported.

While this screen cannot report all possible interactions that might occur under some particular experimental condition, it does provide a great deal of focus on which proteins should be further studied for their implications in cellular processes and networks, and greatly reduces the number of pairwise interactions that might be necessary to studey absent the new evidence.

## Photoswitched fluorophore methods

Fluorescent markers, either externally introduced, or genetically expressed are even more useful when they can be conditionally switched on and off or when the emission spectra can be conditionally modified. These switched fluorophores can be used like other fluorophores for traditional fluorescent imaging; however, they also have application in following the dynamics and translocation of the protein fused to the fluorophore (similar to methods that use bleaching). Further, they have enable new super-resolution microscopy methods, which will be discussed after first considering how translocation and cellular diffusion can be studied with proteins like photoswitched cyan fluorescent protein, PS-CFP (Chudakov, Verkhusha et al. 2004).

PS-CFP is a mutated version of aceGFP (a colorless monomeric molecule from *Aequorea coerulescens* (Gurskaya and al 2003)). Like other photo-switched fluorophores, the fluorescent capability of PS-CFP can be enabled and reset by prior exposure to certain wavelengths. However, alternate exposures can cause a permanent shift in its emission spectra. Initially, PS-CFP has an excitation peak at 402 nm with a peak emission at 468 nm; however, stronger exposure (5-10 W/cm2 for a few seconds with a violet laser, 404 nm) cause an irreversible molecular change shifting the excitation peak to 490 nm and the emission maximum to 511 nm.

By exploiting the irreversible spectral shift, it is possible to illuminate and change the emission color of just the proteins within a small region. These proteins can then be tracked by their different color to make otherwise difficult or impossible measurements involving transport and other motions. For example, Chudakov (2004) tracked parts of two thin filopodia regions in in HEK293 cells which achieved a speed of 0.021-0.024 um/sec. The same methods can be used to track molecular transport during signaling, metabolism, and cell death.

Photo-switched fluorophores can, thus, be used like traditional fluorophores or, as above, to discover dynamic information in living cells. However, one of their most interesting applications is in ‘super-resolution’ microscopy, which allows imaging below the diffraction limit, i.e., the resolution of features much smaller than 0.2 um, the traditional domain of electron microscopy (EM).

Because EM and immunostaining with attached gold beads (Bozzola and Russell 1991), is a relatively slow and expensive process subject to many experimental difficulties, these new optical methods are an exciting development, see the recent review (Hell 2007). Some of these super-resolution methods use photo-switched fluorophores, which can be switched on and off using techniques to ensure that the imaged area is unlikely to include more that a single active fluorophore, knowledge that enables super-resolution of that single molecule by means of repeated measurements.

Fluorescent, super-resolution methods are a very active area of research, but currently three methods are competing for prominence — Stimulated Emission Depletion (STED) microscopy  (Weiss 2000); Photo Activated Location Microscopy (PALM) (Betzig, Patterson et al. 2006); and Stochastic Optical Reconstruction Microscopy (STORM) (Rust, Bates et al. 2006). The physical phenomena relevant to imaging individual (single) molecules, and details about the physical principles behind STED, PALM and STORM microscopy are reviewed in (Davidson 2007).

Each of these methods achieves resolutions of a few tens of nanometers, and should achieve even better resolutions in the future. Importantly, readily available fluorescent stains (Cy3/Cy5) have been shown to have the required optical switching properties (Bates, Blosser et al. 2005; Heilemann, Margeat et al. 2005); so, new exotic chemistries will not be required to collect the images. Even more recently, techniques for multicolor super-resolution imaging has been demonstrated for two different approaches (Bates, Huang et al. 2007; Donnert, Keller et al. 2007).

These revolutionary, new methods are critical for discovering the detailed molecular interactions required for correct modeling the critical interactions of cellular geometry with molecular interactions. They also provide new ways to calibrate models and provide insight into the molecular dynamics and structural changes induced by cellular processes. However, their full potential can only be achieved when automated to allow massive scans and other high throughput studies.

## *Automation for high throughput laboratories*

Several recent breakthroughs are, indeed, enabling new approaches that can identify collocated proteins within their cellular compartments. A particularly interesting example is the work in Germany (Schubert, Bonnekoh et al. 2006) where a laboratory with sixteen automated microscopes has been developed as a step along the pathway toward high throughput collocalization measurements. In their approach, robots implement serial fluorescent immunostaining using flow cells on the microscope stages. They have demonstrated the ability to sequentially stain and image a large number of different epitopes within a tissue sample (or within individual cells under higher magnifications).

They report the ability to regularly use around a hundred different antibodies, and have shown that various permutations of staining sequences gave similar results. Importantly, they are working to extend these methods to make measurements with a thousand different antibodies.

Optically based, high throughput proteomics systems with automated liquid sampling for the chemistries can be seen as just a machine tool for gathering the locations of hundreds if not thousands of proteins localized to their cellular organelles. Seen this way, it is clear how to scale the process: combine and replicate a single common building block in a massively parallel way. Scaling to a production facility with hundreds of these basic building blocks is very much like any other manufacturing environment with many replicas of the basic machine tools. Like other manufacturing facilities, one must be concerned with the materials coming into the plant, the products being shipped, and the wastes being generated. With respect to that production stream, micro-machined flow cells should give greater experimental control while minimizing the required chemicals and resulting waste.

The application of optically accessible, micro-machined reaction vessels and microscopy is already yielding important new experimental methods, including micro bioreactors (Balagadde, You et al. 2005; Hua, Xia et al. 2006). In addition to methods for manipulating and imaging live cells, there are new opportunities for working with fixed cells. Consider, for example, a time course experiment where, after exposing groups to a treatment, responses are collected every few minutes. The exposed cells could be sorted and moved into channels where they would be rapidly fixed at the specified time points. Once fixed, the automated methods discussed above could be used to identify proteins collocating under the experimental conditions. By conducting as much of the experiment as possible within the micro-machined environment, human errors and problems with reproducibility should be greatly minimized.

Micro-fluidic devices are not restricted to simple uses such as cell sorting, and reagent treatments. They have, also, been deployed in sensor and synthesis application (Lee, Sui et al. 2005). As a result, they may play an essential role in preparing cells for testing signaling pathways. They will also be important for understanding how cells lose and regain homeostasis after mechanical and chemical disruptions. The combination of system identification techniques with automated micro-vessels and optical observations of fluorescently tagged proteins can clarify system dynamics and can discover new regulatory elements, which will be discussed next.

## *Use of the measurements and cell models*

Accurate cell models require the ability to localize proteins very accurately (which super-resolution methods are now achieving), the ability to make measurements about intra-cellular motions and dynamics (some of which can be addressed with methods like the PS-CFP experiments by  Chudakov (2004)), and some way to identify regulation and dynamics within pathways. A recent paper serves as an example of this synthesis of micro-machined reaction vessels under dynamic control, and with optical read-out using fluorescent molecules; all used to for model identification and the discover of a new method of regulation of the galactose utilization network (Bennett, Pang et al. 2008; Ingolia and Weissman 2008).

In this work, Bennet et al. (2008) use system identification methods and an automated micro-fluidic device to study how yeast cells switch back and forth between glucose galactose metabolism. Because glucose is the preferred carbon source, the enzymes required to use galactose are tightly repressed in the presence of glucose. The details of this repression has been well studied and, hence, offer a good test case for applying mathematical models of dynamic systems using experimental data from the micro-fluidic chamber.

Briefly, they constructed a micro-fluidics device with individual reservoirs for glucose and for galactose. The device allowed either glucose or galactose to flow through a 4 um tall growth chamber confining *S. cerevisiae* cells that carried a GAL1 fusion with a fluorescent reporter protein. Hence, they could dynamically change the concentrations of glucose and galactose while observing the intensity of GAL1 expression, an measure of the dynamic switching between glucose and galactose metabolism. With computer control of both the micro-fluidic device and the microscope, they collected the response to sinusoidal osculation in relative concentration at different frequencies. Importantly, they discovered that the degradation of some galactose enzymes depended on the concentration of glucose, a previously unknown element of the regulation. After including a term that captures the degradation of mRNA by glucose, the new model accurately reproduces the observed dynamics.

This experiment demonstrates a method that can be expected to become widely used in studying cellular responses. The new experimental method is enabled by advances in micro-scale reaction chamber, optical proteomics, and application of methods for studying dynamical systems that have previously been developed in other fields. It is an example of an automated experiment, but the required automation is applied just at the moment of experimentation.  In many cases, the cellular behavior to be studied will be much more complex than glucose/galactose metabolism switches, and the genomic manipulations required to have optical reports could be extensive. To allow rapid experiments, these manipulation must also be automated.

However, just automating steps that can be done by hand is not a complete answer because the combinatorial complexity is such that millions of strains could be needed. The maintenance of these strains would be costly, and very error prone (even with automation). Instead, the long range goal might be some sort of ‘just in time’ strain synthesis, so that it is so inexpensive and quick to make a modified strain, that it is not necessary to maintain stored libraries. As these methods develop the boundary between *in vitro* and *in silico* will blur and experimentation should become very high throughput with vanishingly small costs for individual experiments. Cell modeling will not only benefit from the experimental results, but will be deeply embedded in the apparatus and experimental analysis, itself.

At the moment, many possibilities are constrained by the required physical work. For instance to create a new fusion protein requires the design of primers specifically for that protein’s gene and the fluorescent reporter, followed by a great deal of bench-scale work. And if a different report, or something like KillerRed, or a protein fragment complementation screen is required, most of the work must be redone.  As a step away from this ‘re-do it every time’ approach this project has developed a chimeric construct that will allow the insertion of a photoswitched cyan fluorescent protein in place of GFP in any of the unique strains in the yeast GFP set.

Because GFP is not suitable for most of the super-resolution methods, the new ability to have photo-switched strains opens the door to very accurate protein tracking and collocations using molecules like PS-CFP2, Dronpa, and Kaede.

These GFP strains contain the same sequences specifying the GFP protein and the selection marker; so, it has been possible to create a new genomic construct to replace an individual GFP gene with the new PS-CFP gene. Further, because this construct exploits similarities introduced behind the gene of interest by the creation of the original strains, the new construct can be used, without change, to modify each and every one of the existing strains.

The top frame in Figure 1 shows how the existing GFP strains have had the stop codon (for the strain-particular protein) replaced by the gene for GFP, and the required terminator sequence (from the alcohol dehydrogenase gene, in this case). The construct also carries the gene for a selection marker (in this case the base strain required supplemental histidine, and this selection marker gives the transformants the ability to make histidine).

The bottom of the figure show how a new construct matching the existing strains with about 40 nucleotides at each end is integrated into the yeast chromosome by homologous recombination. This new insert carries a different selection marker, a gene conferring resistance to kanamycin, which allows identification of successful transformants that are able to grow on Kan+ and His+ plates.

The particular parts of this new construct were not available, so portions from two different plasmids were combined to make a new chimeric construct. In this technique (Pont-Kingdon 2003), a small adaptor, or chimeric primer is designed to match the 3’ end of one piece of DNA and the 5’ end of the other piece. As shown, this primer initially extends the reverse strand of the PS-CFP2 gene. When that strand is available, the F2 primer extends a new forward strand that includes the required 39 nucleotides of upstream homology with the yeast GFP strain and the tiny bit of DNA homologous to the other half of the chimeric fusion (here the new selection marker for kanamycin resistance). That forward strand is able to serve as a forward primer for the selection marker gene, and once extended to the end of the forward strand of the selection marker, there is a complete forward strand of the desired construct. At this point, for the first time, the two outside primers can jointly cause the exponential expansion of the desired chimeric construct.

This PCR product was verified by gel electrophoresis and can now be used either directly from the raw PCR product, or after further amplification in an *E. coli* plasmid, to make the DNA required for transforming any of the GFP strains, as shown at the bottom of the figure.

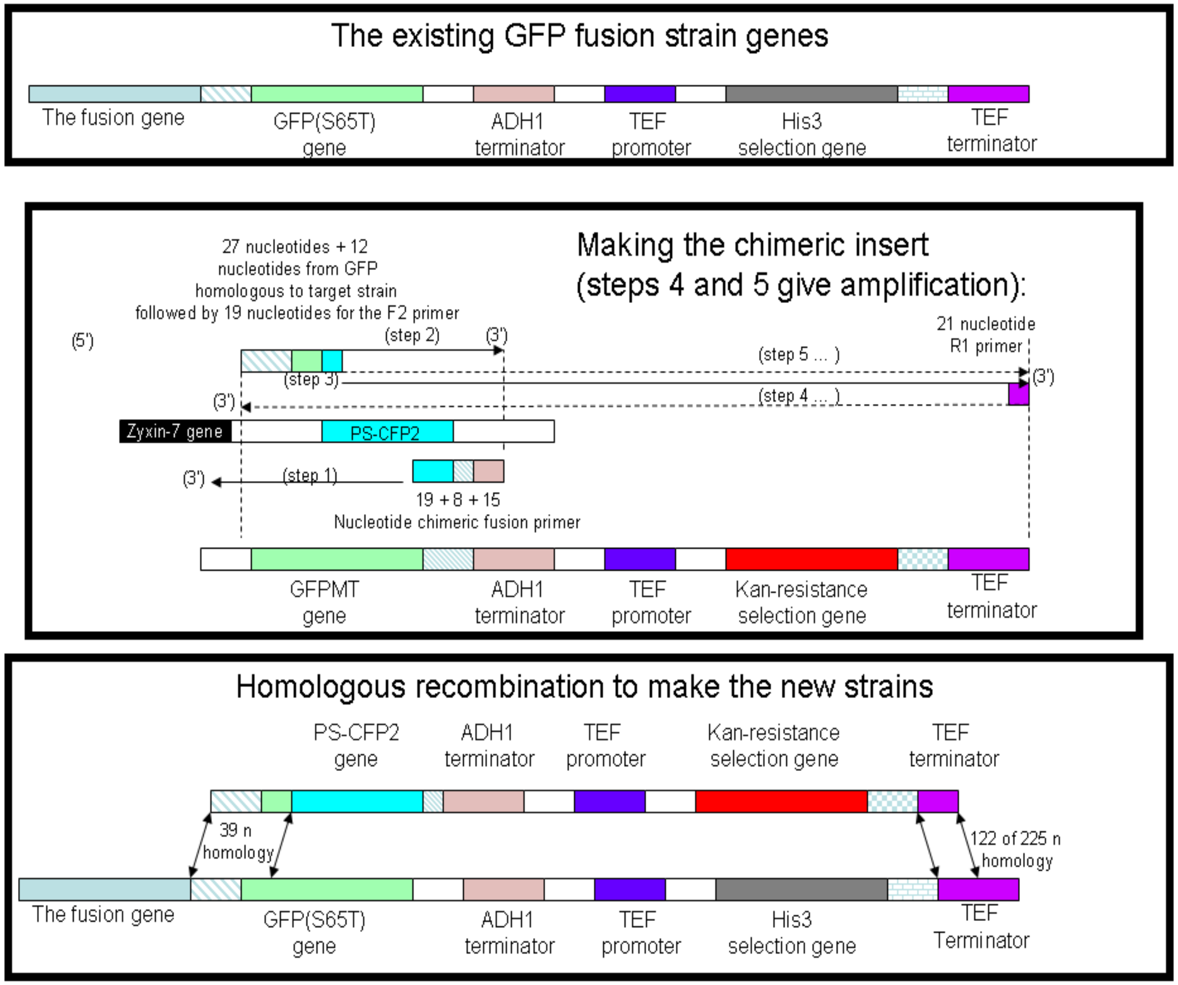
The existing GFP fusion strain genes
The fusion gene
GFP(S65T) gene
ADH1 terminator
TEF promoter
His3 selection gene
TEF terminator
Making the chimeric insert (steps 4 and 5 give amplification):
27 nucleotides + 12 nucleotides from GFP homologous to target strain followed by 19 nucleotides for the F2 primer
(5')
(step 2)
(3')
(step 5 ... )
21 nucleotide R1 primer
(3')
(step 3)
(step 4 ... )
(3')
Zyxin-7 gene
PS-CFP2
(3')
(step 1)
19 + 8 + 15
Nucleotide chimeric fusion primer
GFPMT gene
ADH1 terminator
TEF promoter
Kan-resistance selection gene
TEF terminator
Homologous recombination to make the new strains
PS-CFP2 gene
ADH1 terminator
TEF promoter
Kan-resistance selection gene
TEF terminator
39 n homology
122 of 225 n homology
The fusion gene
GFP(S65T) gene
ADH1 terminator
TEF promoter
His3 selection gene
TEF Terminator

**Figure 1 Construction of a chimeric replacement for the GFP and His3 selection marker in each of the GFP strains, which are shown at the top. The new construct combining the photo-switched cyan fluorescent protein PS-CFP2 and a kanamycin resistance module is created by chimeric PCR, where the first step uses the chimeric primer to extend a copy of the PS-CFP2 gene that will include (at its 5' end) homology to the alcohol dehydrogenase terminator. In step two the forward, outside primer F2 is extended to make a strand carrying a section of homologous DNA for the eventual transformation into the GFP strains, and ending with homology to the selection module. Step 3 has this new stand serving as a forward primer for the selection marker, which once fully extended, can be amplified exponentially by the two outside primers to give appreciable amounts of the required chimeric construct.**

## Discussion and recommendations

The previous sections have reviewed a set of existing technologies that can be combined to create a new research capability, the would allow the analysis of research questions currently out of reach. Sandia National Laboratories has strong capabilities that could contribute to, and extend the capabilities of these methods and instruments. Equally important, our work could greatly benefit if these capabilities were readily available. In particular, our efforts to engineer organisms for biofuels could be accelerated with the adoption of high throughput experiments, and direct (optical) measurements and/or manipulation of relevant molecular pathways via CALI. Fortunately, yeast (*S. cerevisiae*), a critical organism in biofuels research, is one of the best studied model organisms and the tools discussed above were developed for, or are available for use with yeast. In particular, to push these cells far beyond their natural operating range will, likely, require manipulation of apoptosis and stress pathways invoked by extreme conditions.

Interestingly, other work at Sandia National Laboratories investigating responses and mitigations of infectious diseases intersects with those same pathways, and one of the critical enzymes, Akt/PKB (or the yeast homolog, Sch9), is critically involved in stress and apoptosis. By developing high throughput methods and applying Sandia’s expertise in optics, computing, robotics, and micromachines, Sandia National Laboratories could study these critical pathways to advance two of our areas of biological interest (biofuels and infectious disease signatures). Once we have these unique capabilities SNL would be in a position to attract strong research partners and additional funding.

An initial step forward would be to organize a workshop where the needs and relevant SNL capabilities could be explored. The object of such a workshop would be to draw together the internal and external researchers for a possible Grand Challenge proposal or, at least, identify how the many existing and diverse research threads at Sandia might apply these methods, which could become a foundation for internal and external research proposals.

## Acknowledgements

I’d like to thank Roberto Rebeil and his laboratory, especially Bryce Ricken and Kevin Crown for their help and advice during the construction of the PS-CFP2-KanR insert. Also, I am particularly appreciative of the help from Maggie Werner-Washburne and her laboratory who shared the kanamycin resistance plasmid, and who are helping transform the new construct into the GFP strains. Certainly, none of that work would have been possible without the generous gift of plasmids from Hari Schroff and Mike Davidson or the original work in constructing the yeast plasmids by Peter Philippsen. Finally, I’d like to acknowledge the late Ed Oliver, who continuously encouraged me to think about transformative ideas and technologies; I miss him.

# References

Aebersold, R. and M. Mann (2003). "Mass spectrometry-based proteomics." Nature **422**: 198-207.

Ando, R., H. Hama, et al. (2002). "An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein." Proceedings of the National Academy of Sciences of the United States of America **99**(20): 12651-12656.

Balagadde, F. K., L. C. You, et al. (2005). "Long-term monitoring of bacteria undergoing programmed population control in a microchemostat." Science **309**(5731): 137-140.

Bates, M., T. R. Blosser, et al. (2005). "Short-range spectroscopic ruler based on a single-molecule optical switch." Physical Review Letters **94**(10): 108101.

Bates, M., B. Huang, et al. (2007). "Multicolor super-resolution imaging with photo-switchable fluorescent probes." Science **317**(5845): 1749-1753.

Belousov, V. V., A. F. Fradkov, et al. (2006). "LOOK THIS UP AND GET THE PAPER." Nature Methods **3**: 281-286.

Bennett, M. R., W. L. Pang, et al. (2008). "Metabolic gene regulation in a dynamically changing environment." Nature **454**(doi:10.1038/nature07211): 1119-1122.

Betzig, E., G. H. Patterson, et al. (2006). "Imaging intracellular fluorescent proteins at nanometer resolution." Science **313**(5793): 1642-1645.

Bozzola, J. J. and L. D. Russell (1991). Electron microscopy: principles and techniques for biologists. Boston, Mass., Jones and Bartlett Publishers.

Brown, W. L., R. A. Mastico, et al. (2002). "RNA bacteriophage capsid-mediated drug delivery and epitope presentation." Intervirology **45**(4-6): 371-380.

Brown, W. M. and G. S. Davidson (2007). Interactomes to Biological Phase Space: a call to begin thinking at a new level in computational biology. Albuquerque, New Mexico, Sandia National Laboratories.

Brown, W. M. and G. S. Davidson (2007). Interactomes to Biological Phase Space: a call to begin thinking at a new level in computational biology, SAND2007-6278. Albuquerque, N.M., Sandia National Laboratories.

Bulina, M. E., D. M. Chadakov, et al. (2006). "A genetically encoded photosensitizer." Nature Biotechnology **24**(1): 95-99.

Chackerian, B. (2007). "Virus-like particles-flexible platforms for vaccine development." Expert Review of Vaccines **6**(3): 381-390.

Cherry, J. M., C. Ball, et al. (1997). "Genetic and physical maps of *Saccharomyces cerevisiae*." Nature **387**(6632 Supplement): 67-73.

Chudakov, D. M., S. Lukyanov, et al. (2007). "Tracking intracellular protein movements using photoswitchable fluorescent proteins PS-CFP2 and Dendra2." Nature Protocols **2**(8): 2024-2032.

Chudakov, D. M., V. V. Verkhusha, et al. (2004). "Photoswitchable cyan fluorescent protein for protein tracking." Nature Biotechnology **22**(11): 1435-1439.

Colland, F., X. Jacq, et al. (2004). "Functional proteomics mapping of a human signaling pathway." Genome Research **14**(7): 1324-1332.

Davidson, G. S. (2007). Exploiting Single-molecule Imaging in a High-throughput facility to Identify Collocating Proteins *In Vivo*, SAND-2007-8129P. Albuquerque, NM, Sandia National Laboratories.

Donnert, G., J. Keller, et al. (2007). "Two-color far-field fluorescence nanoscopy." Biophysical Journal **92**(8): L67-L69.

Edwards, A. M., B. Kus, et al. (2002). "Bridging structural biology and genomics: assessing protein interaction data with known complexes." Trends in Genetics **18**(10): 529-536.

Förster, T. (1948). Ann. Phys **2**: 55-75.

Giot, L., J. S. Bader, et al. (2003). "A protein interaction map of Drosophila melanogaster." Science **302**(5651): 1727-1736.

Gurskaya, N. G. and e. al (2003). "A colourless GFP homologue from the non-fluorescent hydromedusa *Aequorea coerulescens* and its fluorecent mutants." Biochem. J. **373**: 403-408.

Heilemann, M., E. Margeat, et al. (2005). "Carbocyanine dyes as efficient reversible single-molecule optical switch." Journal of the American Chemical Society **127**(11): 3801-3806.

Hell, S. W. (2007). "Far-field optical nanoscopy." Science **316**(5828): 1153-1158.

Howson, R., W.-K. Huh, et al. (2005). "Construction, verification and experimental use of two epitope-tagged collections of budding yeast strains." Comp Funct Genom **6**: 2-16.

Hua, Z. S., Y. M. Xia, et al. (2006). "A versatile microreactor platform featuring a chemical-resistant microvalve array for addressable multiplex syntheses and assays." Journal of Micromechanics and Microengineering **16**(8): 1433-1443.

Ingolia, N. T. and J. S. Weissman (2008). "Reverse engineering the cell." Nature **454**: 1059-1062.

Ito, T., T. Chiba, et al. (2001). "A comprehensive two-hybrid analysis to explore the yeast protein interactome." Proceedings of the National Academy of Sciences of the United States of America **98**(8): 4569-4574.

Lee, C. C., G. D. Sui, et al. (2005). "Multistep synthesis of a radiolabeled imaging probe using integrated microfluidics." Science **310**(5755): 1793-1796.

Li, S. M., C. M. Armstrong, et al. (2004). "A map of the interactome network of the metazoan C-elegans." Science **303**(5657): 540-543.

McDonnell, L. A. and R. M. A. Heeren (2007). "Imaging Mass Spectormetry." Mass Spectrometry Reviews **26**: 606-643.

McKusick, V. A., M. J. F. O'Neill, et al. (2008). DiHydrofolate Reductase; DHFR, Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University, Baltimore, MD. MIM Number: {126060}: {2/5/2008}: . World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/

Michnick, S. W., P. H. Ear, et al. (2007). "Universal strategies in research and drug discovery based on protein-fragment complementation assays." Nature Reviews Drug Discovery **6**(7): 569-582.

Patterson, G. H. and J. Lippincott-Schwartz (2002). "A photoactivatable GFP for selective photolabeling of proteins and cells." Science **297**(5588): 1873-1877.

Peabody, D. S. (1997). "Subunit fusion confers tolerance to peptide insertions in a virus coat protein." Archives of Biochemistry and Biophysics **347**(1): 85-92.

Pont-Kingdon, G. (2003). Creation of chimeric junctions, deletions and insertions. PCR Protocols (Methods in Molecular Biology). J. M. S. Bartlett and D. Stirling. Totawa, New Jersey 07512, Humana Press Inc. **226:** 511-515.

Rain, J. C., L. Selig, et al. (2001). "The protein-protein interaction map of Helicobacter pylori." Nature **409**(6817): 211-215.

Rual, J. F., K. Venkatesan, et al. (2005). "Towards a proteome-scale map of the human protein-protein interaction network." Nature **437**(7062): 1173-1178.

Rust, M. J., M. Bates, et al. (2006). "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)." Nature Methods **3**(10): 793-795.

Schubert, W., B. Bonnekoh, et al. (2006). "Analyzing proteome topology and function by automated multidimensional fluorescence microscopy." Nature Biotechnology **24**(10): 1270-1278.

SGD (2008). How many genes are there in yeast?, *Saccharomyces* Genome Database (http://www.yeastgenome.org/SGD-FAQ.html) (accessed 20 Aug 2008).

SGD (2008). SGD Search Result for Keyword Mitochondria, *Saccharomyces* Genome Database (http://db.yeastgenome.org/cgi-bin/search/textSearch.pl?query=mitochondria&type=headline) (accessed 20 Aug 2008).

Souslova, E. A. and D. M. Chudakov (2007). "Genetically encoded intracellular sensors based on fluorescent proteins." Biochemistry (Moscow) **72**(7): 683-697.

Surrey, T., M. B. Elowitz, et al. (1998). "Chromophore-assisted light inactivation and self-organization of microtubules and motors." Proceedings of the National Academy of Sciences of the United States of America **95**(8): 4293-4298.

Tarassov, K., V. Messier, et al. (2008). "An in vivo map of the yeast protein interactome." Science **320**(5882): 1465-1470.

Uetz, P., L. Giot, et al. (2000). "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae." Nature **403**(6770): 623-627.

van Meerten, D., R. C. L. Olsthoorn, et al. (2001). "Peptide display on live MS2 phage: restrictions at the RNA genome level." Journal of General Virology **82**: 1797-1805.

Vidal, M. (2005). "Interactome modeling." Febs Letters **579**(8): 1834-1838.

von Mering, C., R. Krause, et al. (2002). "Comparative assessment of large-scale data sets of protein-protein interactions." Nature **417**(6887): 399-403.

Wagner, A. (2005). Robustness and Evolvability in Living Systems. Princeton, NJ, Princeton University Press.

Weiss, S. (2000). "Shattering the diffraction limit of light: A revolution in fluorescence microscopy?" Proceedings of the National Academy of Sciences of the United States of America **97**(16): 8747-8749.

Distribution:

| | | |
|---|---|---|
| 1 | MS-0321 | James Peery, 01400 (electronic copy) |
| 1 | MS-1316 | Mark D. Rintoul, 01412 (electronic copy) |
| 1 | MS-0123 | Donna L. Chavez, 01011 (electronic copy) |
| 1 | MS-9054 | Glenn D. Kubiak, 08300 (electronic copy) |
| 1 | MS 0899 | Technical Library, 09536 (electronic copy) |