

# Genomics:GTL Facility Workshop Report

<http://DOEGenomesToLife.org>

Atlanta, Georgia, June 17-18, 2003

## GTL Facility for Characterization and Imaging of Protein Complexes

Organizer: Michelle Buchanan, Oak Ridge National Laboratory

### Executive Summary

The purpose of this workshop was to define a roadmap for one of the proposed Genomics:GTL facilities, specifically the Facility for Characterization and Imaging of Protein Complexes (referred to in this report as the "Protein Complex Facility"). Over 70 scientists from academic, industrial, and national laboratories attended this workshop, with the majority being life scientists from academic laboratories, who will be the principle stakeholders in the GTL facilities. The workshop was organized in four parts. The first part provided an overview of the Genomics:GTL program of the Department of Energy (DOE) Office of Biological and Environmental Research (BER) and four proposed facili-

ties that will support this research program. These four include protein production and characterization, whole proteome analysis, analysis and modeling of cellular systems, and, the subject of this workshop, characterization and imaging of protein complexes.

The second part of the workshop included talks from experts in approaches for isolating, identifying, and characterizing protein complexes and the analytical and computational tools associated with these processes. The third part involved breakout groups that discussed specific tasks and capabilities that should be incorporated within this facility. These breakout groups allowed input from all participants to define the needs of the biological community and capabilities that should be included as part of the proposed facility.

Attendees strongly supported DOE's vision of establishing this facility as one of the proposed four, specifically focused on providing high-validity data on the identification and characterization of the complete suite of protein complexes in microbial cells. The consensus was that this type of high quality data can be generated only by a focused, high-throughput effort similar to today's large gene sequencing entities. Workshop participants expect this facility to provide information that will profoundly impact both fundamental and applied biological research. Having key data for use by the broad biological community was viewed as a catalyst in harnessing biochemical processes used by microbes to meet

### Contents

Executive Summary .....	1
Overview .....	2
Breakout Group: Growth and Expression .....	5
Breakout Group: High-Throughput Isolation and Mass Spectrometry Identification .....	8
Breakout Group: Imaging and Analytical Tools for Validation and Biophysical Characterization of Protein Complexes .....	13
Summary .....	16
Agenda .....	19
Letter to Confirmed Participants .....	20
Attendees .....	22

DOE's mission needs in energy production, environment restoration, and carbon management.

## Overview

The workshop was focused on defining a roadmap for the Facility for Characterization and Imaging of Molecular Machines. This facility will identify molecular machines; characterize the nature of their intermolecular interactions; and model, predict, and simulate their structure, activity, and dynamic behavior. It will bring together a unique set of state-of-the-art analytical and computational tools to enable for the first time the high-throughput identification and characterization of thousands of molecular machines. Its primary products, high-fidelity data, and associated computational tools will be made available to the broad biological community. These products will provide the basis for developing a fundamental understanding of cellular function. Overall, the 10-year vision is for the facility to be a high-throughput enterprise that produces high-quality data, reagents, and analytical and computational tools for the annual characterization of up to thousands of protein complexes. Specifically, the facility's purpose is to

1. Discover and define the complete inventory of protein complexes in a microbe.
2. Analyze the structure and predict the function of these molecular machines.
3. Develop principles, theory, and predictive models for the structure, function, assembly, and disassembly of multiprotein complexes.
4. Provide high-fidelity data and tools to the greater biological community.

The first step considered by workshop attendees was to define the types of data needed by the biological community to identify and characterize protein complexes. The second step was to define how to provide the highest-validity data that can be used by biologists in their laboratories. Finally, the workshop sought to define how this generated data will be made available to the

greater biological community and what kinds of tools will be needed to access and mine the data produced. Data quality and ease of access need to be as robust and valid as the data we now use for gene sequencing. The following topics were identified as necessary for defining the roadmap for the Protein Complex Facility.

- Facility products
- Critical requirements
- Performance criteria
- Resources needed
- Definition of major "pipelines"
- Alternative approaches
- Timelines
- R&D needed
- Operation model

Participants were charged with formulating a roadmap that includes a comprehensive view of what is needed in this facility, including cell growth, expression of the complex, isolation of complexes from cells, high-throughput identification of complexes, biophysical characterization of complexes, and in vivo and in vitro imaging of complexes. Both short-term—what we can do right now—and long-term needs—what we will need in 5 to 10 years—were to be considered as part of this roadmap. Most important, computational capabilities required to support the facility were carefully considered and integrated throughout all aspects from laboratory information management systems (LIMS) to track samples, systems for data collection and interpretation, databases for data archiving, tools for data mining and manipulation, and algorithms for data modeling, simulation, and prediction.

## Genomics:GTL Facilities in Context

A basic research program, already under way, underlies the GTL program. The proposed facilities will support the larger GTL research program, and this and other workshops will help define their roles in bringing information to the scientific community that will enable basic fundamental science. The facilities are envisioned as enablers of science to

provide knowledge that will eventually turn into applications to meet DOE mission needs. These facilities are expected to come online beginning, under the most optimistic scenario, 5 to 7 years from now. Defining roadmaps for the facilities now will ensure that they will meet the goals of the GTL program and the scientific community.

There are four major goals in the GTL program. The first goal is to identify the molecular machines of life—the protein complexes that conduct the ensemble of processes within a cell. The second goal is to characterize gene regulatory networks. The third goal is to characterize the functional repertoire of natural microbial communities. The fourth goal is to develop computational methods and capabilities to advance understanding of complex biological systems and predict their behavior. Goal 1, which is essentially the basis for the Protein Complex Facility, has three steps that lay the foundation for the GTL program: (1) Identify the complement of protein complexes and their components; (2) Elucidate function and dynamics of complexes—intermediates, nature of interactions, cellular location, and kinetics; (3) Establish how changes arising from such factors as environmental stress and development affect complex formation and function. Goal four, which includes computational tools, also will be a major component of the Protein Complex Facility.

GTL facilities will be tasked with large-scale generation of information that will revolutionize biological research, in a manner similar to the revolution realized during the Human Genome Project by large-scale gene sequencing. In the case of protein complexes, rather than small-scale, single-pathway analysis, an entirely new comprehensive approach must be formulated to the identification and characterization of protein complexes.

### **Relationships Among the Protein Complex Facility and Other GTL Facilities**

Each of the four proposed facilities is designed to focus on a specific aspect of the GTL program, with the facilities closely interacting as a combined resource to the GTL community. The Protein Production and Characterization Facility focuses on the production of proteins, affinity re-

agents, and other reagents that will be needed by the community and other facilities. In the case of the Protein Complex Facility, the Protein Production and Characterization Facility will provide affinity reagents (antibodies and tagged proteins) to isolate protein complexes, protein standards, and stable isotope-labeled proteins for mass spectrometry (MS), neutron scattering, and other techniques. The Whole Proteome Facility will provide valuable data on the complement of proteins and metabolites in a microbial cell. Data from the Protein Complex Facility will be required for the facility on Analysis and Modeling of Cellular Systems. In addition, the Protein Complex Facility will provide information to the Protein Production and Characterization and Whole Proteome facilities as well as direction for future studies. Synergies among these four facilities will provide valuable data, reagents, and resources to the greater biological community.

### **Significance of Protein Complexes**

Proteins rarely act alone; rather, they are organized and carry out all the internal biological processes by molecular contact. These biologically important protein-protein interactions (as well as protein-RNA, protein-DNA, and other biomolecular complexes) modify and dictate specific protein states and, in turn, define cellular states. Thus protein complexes literally are the workhorses of the cell. Protein complexes are key components in various biochemical pathways occurring in a cell and are thus a central component of systems biology. Knowledge of the complete set of protein complexes in a microbial cell and their dynamics will enable us to elucidate these biochemical pathways that define biological function. The collection of protein complexes does not have just a simple linear relationship—reactants A and B react to give product C—but are involved in a large, dynamic network. Once these complex relationships are understood, we can establish how pathways are altered under various environmental conditions, how they differ from one organism to another, and how microbes interact with each other in a microbial community. For example, specific pathways that will enhance hydrogen generation can be turned on or off in an organism.

Defining protein complexes was viewed as a critical step in defining a roadmap for this facility. Many types of protein complexes exist in cells. Many are assembly intermediates, while others are fully functional molecular machinery. For example, key cellular information comes from multienzyme complexes that can result in increased reaction rates, reduced side reactions, and direct transfer of metabolites. Many protein complexes are literally machines—those that actually do work (e.g., folding and motors). Others are so-called array machines, such as light-harvesting systems, ribosomes, and others, that carry out complex conversions and do the vital work of many organisms. Complexes also can be defined

from an operational perspective from subcellular fractionation: stable/soluble, transient/soluble [low affinity, short lifetime], and membrane.

The workshop was broken into four major breakout sessions: Growth and Expression, High-Throughput Isolation and Mass Spectrometry Identification, Imaging and Analytical Tools for Validation and Biophysical Characterization of Protein Complexes, and Bioinformatics and Computation. The findings of these four breakout sessions are outlined below.

# Growth and Expression

*Participants: Mike Adams (U. Georgia), Carl Anderson (BNL), Fred Brockman (PNNL), Brian Davison (ORNL), Terry Hazen (LBNL), Bob Kelly (North Carolina State), Biswarup Mukhopadhyay (Va. Tech), Dale Pelletier (ORNL), Bob Tabita (Ohio State), and Frank Larimer (ORNL)*

## Challenge for Growth and Expression of Microbial Protein Complexes

The main challenge is to generate various quantities of protein complexes (possibly millions of different proteins) from a variety of species (potentially tens of thousands of gene sequences). To meet the goals of the GTL program, multiple different species of microorganisms must be grown in high throughput under a variety of carefully controlled-state conditions. In addition, recombinant clones will require culture under controlled conditions with high expression, high throughput, and production turnaround. This requires technology to express intact protein complexes in wild-type and recombinant cultures under well-characterized conditions so that the molecular machines may be separated and analyzed.

For the purposes of this facility, a "controlled-growth" state is a sample of microorganisms under a determined growth condition. Conditions to be controlled must include environment (temperature, pH, media, substrate, light/dark, oxygen); growth state (exponential, steady-state, balanced, stationary); operation (batch, continuous); and harvest (age, lag, concentration, handling conditions). The microbial state is known to influence the expression and activity of protein complexes. It is an open question of the "natural" variability of states within a controlled-growth microbial population. Variability of expression under nonstandard conditions also is poorly understood and remains an important research need.

Microbial growth and expression specific to the identification and characterization of protein complexes have two central questions: (1) What is the sample size needed for analysis? and (2) What is the method of protein complex isolation? The required sample size is expected to decrease with improvements in analysis and visualization. Current estimates are that about a gram of wet microbial cell biomass will be required for the majority of MS-based analyses prior to isolation of the complexes. However, 1 to 10 mg of a partially purified complex currently is needed for biophysical characterization. The number of cultures and clones to be grown is quite dependent on the strategy of protein complex isolation. If affinity reagents are readily available for each protein within the complex, then a smaller number of wild-type cultures can be grown under controlled conditions. This biomass sample then will be used for multiple pulldowns with each target affinity reagent. However, if the affinity reagents are not readily available and tagged proteins expressed in recombinant clones of the microbe of interest must be used, the number of individual cultivations required increases of 1000-fold (one cultivation for each clone for each tagged protein). The workshop group expressed confidence that the first approach will be realized by the Facility for Protein Production and Characterization. However, they felt that the Protein Complex Facility should develop a supplemental strategy capability for growth and isolation of "difficult" protein complexes (membrane and low abundance) based on tagged proteins expressed in the original native microbial hosts. Note that expressing these tagged proteins under controlled-growth states in their native host is important to understanding the complexation with other proteins in

a "normal" state-not in a conventional production microbe like *E. coli* where the rest of the potential protein partners will not exist.

Most discussion was based on this affinity reagent assumption, which restricts growth to 1- to 2-L vessels. If all affinity reagents could be used on "one microbial sample," even larger cultivations may be advisable. Under these conditions, true high-throughput cultivation is unlikely. However, most of the technology for this exists.

The typical simplified "pipeline" for a new microbe would be to grow native or wild-type microorganisms under a reference state, harvest, use affinity reagents to pull down complexes from that native biomass sample, and then analyze complexes by MS, imaging, and other analytical tools. This process would be repeated under different growth states to extend coverage. The Whole Proteome Analysis Facility would establish reference growth states for specific organisms.

One identified research need in growth and expression is to pilot integrated setups for growth, expression, and harvest of microorganisms. The large pilots should include several model organisms across the range of metabolic and phylogenetic possibilities (target would be ten microbes). These pilots would test integrated equipment, growth-data analysis and capture, harvest protocols, and sample storage. Variability within and between samples would need to be assessed to determine the number of replicates required to minimize variations due to experimental conditions and to the multistate of single microbes within reproducible community. Techniques should be assessed to provide cultivations with well-measured growth states, including all conventional measurements (pH, temperature, dissolved oxygen, and more) as well as off-gases and analysis of metabolites (e.g., HPLC for sugar). Newer analytical procedures such as protein chips and mRNA expression arrays should be brought online.

Unlike a typical industrial fermentation facility, this growth and expression facility will be required to have conventional and unusual fermenters. A conventional fermenter operates

under batch cultivation, well mixed, with air bubbled through to provide oxygen and growth on some carbon source. This facility will require such additional equipment as chemostats, photobioreactors, and biofilms to address a wide range of organisms and growth conditions (anaerobic, microaerobic, and extremes of pH, temperature, and salt). Improved and alternative harvesting techniques are needed for chemostatic cultivation and for sample stabilization of grams of biomass. One particularly critical opportunity provided by this range of cultivation equipment is the potential to discover unknown complexes by stressing microbes into novel environments.

## Research Needs in Growth and Expression

One research need is to reduce well-controlled cultivations to the milliliter scale to allow greater numbers of variants to be grown. Current 96-well plate technology does not allow optimal control, expression, or analysis for two reasons: anticipated improvements in isolation and analysis methods to decrease sample size and a backup approach when affinity reagents are not available. Another need to increase the range of cultivations is by scale-up to larger well-controlled cultivations to reduce effort and examine more novel conditions. We need to push growth conditions and methods to assess protein complex function. Ultimately, this facility should be able to perform the equivalent of "culturing the unculturable." This will require analysis of direct environmental samples for complexes, growth of mixed cultures, and simulated experiments to mimic environmental samples.

There are specific research needs for imaging complexes, initially for better controlled growth states within current and developing imaging devices (i.e., maintaining oxygen, substrate, and pH). Ultimately this path might lead to single-cell controlled state reactors with imaging and complex analysis and identification. On another long-term path, adapting the imaging equipment to operate directly in the instrumented fermenters would be desirable.

In molecular biology, the ability to manipulate the target microbes should be developed as part of the GTL program for use at this and other facilities. These manipulations include tag protein expression, knockouts, and other applications. Recombinant tools for manipulation of entire protein complexes (protein-protein, protein-DNA, protein-RNA) in new microbes need to be developed. Specifically, this includes multigene expression for entire packages of complex components. Finally, advanced analytical capabilities will be needed to provide a more complete understanding of growth conditions. Online analysis of metabolites would identify small molecules associated with protein complexes and help fill in cellular network

information. This may be accomplished by use of stable isotope analysis of metabolites.

Computational tools will be used across the growth and expression efforts. Bioinformatics will assist in selecting target proteins and tags to be used in isolating complexes. LIMS needs to be incorporated from the very beginning of these processes to track samples and annotate growth conditions. LIMS also will be used to archive samples in cold storage and freezers. Heuristic computational tools will be needed to interpret and evaluate growth data. Instrumented growth chambers can be controlled more precisely by computer feedback.

# High-Throughput Isolation and Mass Spectrometry Identification

*Participants: Mark Biggin (LBNL), Bob Foote (ORNL), Michael Giddings (North Carolina, Chapel Hill), Andrei Gorin (ORNL), Greg Hurst (ORNL), Steve Kennel (ORNL), Andrew Link (Vanderbilt), Tim Palzkill (Baylor College of Medicine), Karin Rodland (PNNL), Fred Schachat (Duke), Anup Singh (Sandia), Robert Strausberg (TIGR), Nathan Verberkmoes (ORNL/University of Tennessee, Knoxville), Yisong Wang (ORNL), Dong Xu (ORNL), Joe Zhou (ORNL)*

## Isolation of Protein Complexes

Table 1 lists a series of techniques that were identified as current or emerging candidate technologies for isolating protein complexes.

Conventional biochemical isolation techniques, such as sucrose density gradient centrifugation and preparative column chromatography, are widely used at present to isolate particular proteins and complexes. These techniques are trusted and often robust, but the question is whether they can be automated and applied at an organism- or genome-wide scale due to the tailor-made nature of the approach.

A number of techniques based on affinity chromatography were discussed. One affinity-based

technique, currently implemented in a handful of laboratories around the world, is *in vivo* generation of a fusion between each protein encoded by the genome and an affinity "tag." Such fusions can be produced endogenously by introduction of a plasmid into cells of the bacterial species of interest, or exogenously by expression in a well-behaved host such as *E. coli*. In the former case, affinity chromatography using a resin selective for the affinity tag (a "pulldown") should isolate the tagged protein along with its interaction partners, while in the latter case, the tagged reagent protein would be mixed with a lysate of the organism of interest to insert into appropriate complexes before the pulldown. Other affinity-based techniques involve the use of reagents that are highly selective toward the complex of

**Table 1. Protein Complex Isolation Techniques**

Applicable to "stable" complexes and interactions	• "tagless" approach: automated, high-resolution multidimensional chromatography on whole lysates
<ul style="list-style-type: none"><li>• conventional biochemical isolation</li><li>• create fusion proteins with affinity tags<ul style="list-style-type: none"><li>• endogenous</li><li>• exogenous</li></ul></li><li>• create reagents with affinity toward complexes<ul style="list-style-type: none"><li>• antibodies from phage display techniques</li><li>• array formats</li><li>• baits<ul style="list-style-type: none"><li>• proteins</li><li>• lipids</li><li>• nucleic acids</li></ul></li></ul></li></ul>	Applicable to "transient" complexes and interactions
	<ul style="list-style-type: none"><li>• chemical cross linking to stabilize complexes</li><li>• imaging</li><li>• yeast two-hybrid</li><li>• substrate-trapping mutant enzymes</li></ul>



interest. Antibodies are one such reagent and could be generated by a variety of techniques. Other affinity reagents include nucleic acids (SELEX) and lipids or lipid vesicles. These types of reagents are all amenable to use in large-scale arrayed formats. The present rate-limiting step is difficulty in generating such affinity reagents for "all" protein complexes in an organism, although this is an active area of currently funded GTL research. The GTL Facility for Protein Production and Characterization ultimately will be responsible for preparing most affinity reagents for this facility; as noted above, however, the facility will need to include some capabilities for producing the reagents, particularly for membrane and rare complexes.

A "tagless" strategy for isolating protein complexes is also possible. This strategy would use automated, high-resolution multidimensional chromatography to separate cell lysates into many (say 10,000) fractions. The strategy might enable looking in a few days at a large fraction of an organism's protein complexes. A further advantage of this approach is that no affinity reagents would have to be produced. A substantial bioinformatics component would be required to produce information on identities of components of complexes based on their chromatographic behavior.

The isolation techniques described above generally would be applicable to more stable complexes. Much relevant knowledge, however, is likely to involve transient interactions among proteins. Techniques discussed that could be applied to these interactions are listed in Table 1. There also is a need to consider various ways to

incorporate microfluidics in the isolation and sample introduction to MS or other downstream analyses. Microfluidics was noted to have particular potential for minimizing sample amount, perhaps even to the level of single-cell analysis. Further, using microfluidic devices to combine affinity sample isolation steps, perhaps with multistage columns to perform pull-downs with sequential affinity reagents, offers the potential to decrease the number of required sample-handling steps that would minimize sample loss and contamination.

A key issue in isolating protein complexes will be to establish sample amounts needed for subsequent characterization and identification techniques. Table 2 gives estimates for current technologies, and advances undoubtedly will result in lower requirements. Widely varying abundances of various complexes produced by the cell also are an important factor in determining starting amounts of biomass.

One recognized problem with the combination of affinity pull-downs and MS is the detection of nonspecifically interacting proteins. Use of multiple isolation strategies, including the "tagless" approach described above that does not rely on affinity purification, might take advantage of complementary selectivities for identifying specific vs nonspecific binding. Computational tools would be critical in sorting through data in these multiple isolation strategies to identify true complexes from artifacts.

**Table 2. Estimated Sample Requirements for Identification and Characterization Techniques**

Technique	Sample Requirement
Mass spectrometry	ng to $\mu\text{g}$
Cryoelectron microscopy	$\mu\text{g}$
Reconstitution of complexes for functional assays	$\mu\text{g}$ to mg
Optical Imaging	variable
AFM	$\mu\text{g}$
Structure (neutron scattering, X-ray scattering, NMR)	mg
Biophysical characterization	variable

## Mass Spectrometry Identification of Protein Complexes

Because of its demonstrated and growing strengths for protein identification, MS undoubtedly is the current top candidate technology for identifying subunits of protein complexes. The continuing rapid growth in MS capabilities make it difficult to foresee what types and numbers of instruments would be most appropriate for the Facility for Whole Proteome Analysis. However, Table 3 gives several different types of instruments that either are widely in use today for these types of applications or show promise for the near future.

Other tools used for identifying protein complex components are interaction arrays (protein chips), yeast two-hybrid techniques, and others; but it was generally felt that MS would be the workhorse tool for identifying protein complexes in the foreseeable future.

**Table 3. Classes of Mass Spectrometry Instrumentation to Consider for the Facility for Whole Proteome Analysis**

LC-MS-MS on QIT platform  
LC-MS or LC-MS-MS on FTICR platform  
MALDI on platform capable of MS/MS, such as TOF-TOF or other

### Abbreviations

- MALDI: Matrix-assisted laser desorption/ionization
- LC: Liquid chromatography
- MS: Mass spectrometry
- MS-MS: Tandem mass spectrometry
- FTICR: Fourier transform ion cyclotron resonance mass spectrometry
- QIT: Quadrupole ion trap
- TOF: Time-of-flight mass spectrometry

## Integrating Isolation and Identification

Although discussed separately above, isolation and identification activities will need to be integrated into various facility workflows. This will minimize sample handling (and possible contamination), reduce the amount of starting material required, and increase sample throughput. Therefore, a careful matching of the throughput of isolation and identification capabilities will be required. One starting point for performing this matching is to define the facility's desired throughput in complexes per year. Ideally, it would be possible to identify "all" complexes in a bacterial system by isolating each protein expressed by the bacterium along with its interaction partners, which could then be identified. Estimating that the genome of a typical bacterium encodes some 5000 genes, one could estimate the scales of various facility operating components. The first set of complications to this simplistic model would be that maybe 1000 of these 5000 isolations and identifications would be "easy" stable, soluble complexes; another 1500 might be more difficult membrane proteins, and another 2500 might represent proteins involved in transient interactions. Variants, posttranslational modification, and expression levels represent further complications. Thus, graded levels of effort would be required.

Methods for scaling isolation of current techniques to higher throughput will require use of 96-well or higher-density format. Current pilot projects are addressing this issue as isolation methods are established. Another route to parallel implementation of isolation techniques might be multiple parallel microfluidic channels or columns that could be manipulated independently based on feedback from a sensor array. Both approaches also will expedite sample throughput, decrease potential sample loss and contamination, and reduce needed sample amounts. In addition, microfluidics holds the promise of single-cell analysis.

## Research Needs for Isolation and Mass Spectrometry Identification

A number of candidate technologies for isolating and identifying protein complexes are being evaluated in current GTL pilot projects. Results generated by these projects and by the facility itself will improve the efficiency of subsequent measurements on, for instance, complexes produced under different growth conditions for a given organism. One possible strategy is to perform an initial characterization of an isolated complex by a technique that provides relatively complete protein identification information, such as LC-MS-MS. Subsequent measurements on this complex, say under different growth conditions, could employ a higher-throughput technique such as 2-color, 2D gel electrophoresis or protein chips. This approach could identify changes in expression levels, variations, and posttranslational modifications that could be targeted for further study, without the expense of exhaustive LC-MS-MS on the entire complex.

While fusion proteins containing affinity tags are currently in use, there is considerable room for improvement in this technology. For example, can new tags be designed that allow more efficient isolation, and are less obstructive to protein function and complex formation? Improvements are needed in methods of introducing affinity tags into microbial species. This challenge will require close consultation with experts on genetic systems for particular microbes.

There also is a critical need to accommodate the entire dynamic range of complexes in cells (i.e., to analyze abundant complexes, complexes present at a single copy per cell, and every level of abundance in between). Novel technologies need to be developed to circumvent current limitations in dynamic range. One possible method for "leveling" the isolated amounts of various complexes is to load an affinity column to saturation with a targeted complex. Abundant complexes will occupy all available binding sites on the column after passage of a small amount of lysate, while a comparable amount of a rare complex can be isolated by passing a larger amount of lysate over the column. Sensitivity also is an issue

for analysis of protein complexes. Ideally, single-cell analysis would be desired, with the ability to detect components present at one copy per cell.

As mentioned above, protein complexes with different properties will require different isolation strategies. The major dividing lines identified were transient vs stable complexes and soluble vs membrane complexes. Membrane complexes could be divided further into integral membrane proteins vs membrane-associated complexes. In addition, different cell compartments have their own membranes and membrane proteins (periplasm, outer membrane) that could be isolated separately. To complicate matters more, identification of characterization techniques to be applied to isolated proteins will dictate limits on the use of certain reagents in the isolation. For example, MS generally is incompatible with many detergents used to solubilize purified membrane proteins. New approaches are needed to identify and characterize membrane and transient protein complexes. Imaging holds considerable promise for observing transient complexes, but, as discussed later, new imaging capabilities will be needed to realize this promise fully.

## Current Gaps in Mass Spectrometry Identification Technologies

At the interface between isolation and identification, the issue of more efficient use of the sample presented to the mass spectrometer needs to be assessed. The number of protein or peptide ions surviving the voyage to detector in a mass spectrometer is substantially lower than the number of molecules presented at the inlet. Research is needed to enhance the number of ions produced from the protein/peptide and efficiently introduce the ions into the mass analyzer. Further, new ionization techniques are needed to enhance uniformity in analyte ionization that will allow improved quantitation and analysis of nonpolar moieties.

Improved relative and absolute quantitative MS methods are needed to allow stoichiometries,

changes in expression levels among growth states, and other important information to be obtained. Current methods all rely on stable isotope labeling to allow the ratios of chemically identical species with different masses to be determined. All current techniques suffer from some limitation such as limited throughput, prohibitive reagent expense, or lack of accuracy.

Much current MS-based protein analysis is performed using the "bottom-up" strategy, in which proteins are first digested enzymatically before MS analysis and identification of the resulting peptides. However, there is growing interest in "top-down" proteomics, in which the molecular masses of intact proteins are measured by MS. A major advantage of the top-down technique is the ability to identify post-translational modifications in proteins, which will be critically important in the analysis of protein complexes. Small changes in protein interactions within a complex can profoundly affect the resulting cellular process. Therefore, both "bottom-up" and "top-down" capabilities should be present in the Facility for Whole Proteome Analysis. However,

a number of technical hurdles remain before the routine and high-throughput application of the top-down strategy can be realized. Many of these bottlenecks can be best addressed by novel bioinformatics approaches. Unlike the analysis of peptides resulting from a "bottom-up" approach, there currently are no databases and bioinformatics tools for the analysis of intact proteins and various modifications.

The past decade has seen astonishing improvements in capabilities for protein analysis by MS. Additional novel approaches are expected in the next few years and should be monitored for possible incorporation into this facility to improve sensitivity, dynamic range, quantitation, and information content of MS. For example, new approaches that use MALDI or variants of this technique may offer potential speed advantages over techniques based around an LC separation. Protein chips (interaction arrays) that can be "read" by MS techniques such as SELDI, MALDI, or electrospray methods also may prove to be valuable technologies for the rapid analysis of protein complexes.

# Imaging and Analytical Tools for Validation and Biophysical Characterization of Protein Complexes

*Participants: Mitch Doktycz (ORNL), Cynthia Peterson (UTK), Leonard Spicer (Duke), Yisong Wang (ORNL), Nathan VerBerkmoss (UTK-ORNL), Bob Strausberg (TIGR), Betty Mansfield (ORNL), Mike Knotek (Consultant), Karin Rodland (PNNL), Frank Harris (ORNL), Jizhong Zhou (ORNL), Bob Tabita (ORNL), Danny Rintoul (SNL), Robert Scott (University of Georgia), Fred Schachat (Duke), Dong Xu (ORNL), Ed Uberbacher (ORNL)*

## Challenge for Growth and Expression of Microbial Protein Complexes

High-throughput analysis of protein complexes by MS will not provide sufficient confirmation of their identities. Additional approaches will be required to validate the presence of the complexes in cells and to understand under what cellular conditions these complexes are present. In addition, data on the dynamics of these complexes are needed to help understand their biological function. Biophysical data, such as binding affinities and interaction interfaces, need to be defined to help elucidate function and the effects of modifications on function of complexes. Thus, the collection of analytical technologies required by this facility is quite broad. Required tools break down into those for imaging and for biophysical characterization. In particular, various fluorescence, scanning probe, and electron microscopy techniques are expected to be cornerstone tools for the facility. Clearly, computationally based modeling and characterization are closely coupled. Rather than partition needs along a technology basis, the group divided tools among the types of required functional information.

## Imaging and Analytical Tools for Biophysical Characterization

These types of required information are broken down into areas including identification, structure and morphology, dynamics, and energetics. Each is discussed below.

**Identification.** Tools for identifying protein complexes-protein-protein, protein-DNA, protein-small biomolecule, or other complexes-will be essential to complement data obtained by mass spectrometry (discussed above). For example, in vivo imaging techniques will be critical for verifying data based on MS. Fluorescent labels introduced into protein components of complexes can be monitored in living cells using optical microscopy to confirm the co-localization of proteins identified by mass spectrometry. Multiple tags can be monitored simultaneously to follow protein positions as a function of time. In addition, this approach will be particularly valuable for identifying weak, transient, and membrane-bound interactions. Scanning probe-based techniques may be particularly useful for identifying complexes associated with membrane-bound proteins. Improvements in imaging techniques based on fluorescence and scanning probes can yield approaches that can meet higher-throughput demands.

Protein and other types of arrays, coupled with effective detection methodologies, will be needed

to confirm specific associations. Various detection methodologies for these arrays could include fluorescence, surface plasmon resonance, force microscopies, calorimetry, MS, and electron microscopy. Many of these techniques can be used to determine if biomolecules are specifically interacting, and some can be applied in either microarray or chromatographic formats. Additionally, hydrodynamic techniques, yeast two-hybrid, and computer-based prediction were identified.

**Structure and Morphology.** In addition to identifying complexes, information on the structure, morphology, and stoichiometry of biomolecular complexes was identified as essential. Imaging techniques can provide such information. Various forms of electron microscopy, tomographic imaging, and scanning probe microscopy could be improved to meet higher-throughput needs of this facility. Scattering techniques, such as neutron scattering, would help to provide insight into the interactions of components in protein complexes. Additionally, several biophysical techniques were highlighted, including light scattering, NMR, and hydrodynamic techniques such as analytical ultracentrifugation.

**Dynamics.** Characterizing the time-dependent features of biomolecular complexes will be necessary for relating functional information to biomolecular complexes. Dynamic measurements down to nano- and picosecond increments will be required, and several technologies are relevant to these measurements. These include fluorescence-based techniques such as fluorescence resonance energy transfer, fluorescence polarization anisotropy, magnetic resonance-based techniques, and advanced optical microscopies. Techniques for performing these measurements on the single molecule and whole cells will be required.

**Energetics.** The thermodynamics associated with biomolecular interactions and complexes also will be needed. This information will be essential for relating biomolecular interactions and function on a physical chemical basis. Several technologies were identified as critical including calorimetry, force microscopy, and laser trapping techniques. Additionally, a range of biophysical

techniques that enable monitoring of thermal- or solvent-induced structural transitions would be desired.

Quality assurance for characterizing protein complexes will require the use of standardized protocols and controlled measurement environments. Further, examination will be required using multiple techniques that provide a common answer consistent with genetic-based information and statistically determined. A related concern is to define what a "real" biomolecular interaction is. This involves distinguishing signal from noise and establishing biological relevance of a complex.

## Research Needs in Imaging and Biophysical Characterization

Numerous technological advancements will be required to realize facility goals. In general, technologies for measuring weak, transient, and membrane-bound interactions are lacking. Sample-throughput issues also were a concern. Although many techniques discussed above are amenable to higher throughput, commercially available instruments that fit these needs are lacking. Another general concern was the availability of samples that meet the needs of biophysical measurements. Higher quality, well-defined samples, and, typically, higher quantities of samples are required for biophysical measurements. High-throughput techniques for producing high-purity samples will be required. In addition to these general concerns, specific identified needs included:

- Advancements in single molecule measurement techniques - Single-molecule techniques are expected to play a large role in characterizing biomolecular complexes. Improved instrumentation, calibration standards, and robust labels are needed.
- Improvements in labeling - Sample tagging techniques and materials play a vital role in many imaging and biophysical characterization techniques. For fluorescence-based techniques, dye with longer lifetimes, resistance to photobleaching, and different emission wavelengths are desired.

- Whole-cell measurement techniques - Methods for characterizing entire cells at the molecular level will be needed. Further development of these tools will be required to define the function of complexes in a live cell.

Long-term technology needs are those that fulfill the goal of connecting genome information to function. Ultimately, the sum of information being produced from this facility should enable the prediction of complexes, physiology, and function based solely on gene sequence. This information should enable the transition to codebases that simplify instructions encoded in databases. High-throughput automation that enables comprehensive assessment of the spatial and temporal characteristics of molecules within cells is dreamt about. Extending these measurements for simultaneously characterizing entire communities of cells also was discussed.

Computational and bioinformatics needs are intricately woven within all aspects of imaging and biophysical characterization of protein complexes. Beyond obvious needs for data handling and interpretation, new tools for molecular biophysics are needed that relate biophysical characteristics to molecular structures, bioinformatics tools that effectively archive and mine biophysical and imaging data, and new systems biology models that integrate the various data types. Computationally guided experiments also will be critical for focusing time-intensive experiments and narrowing parameter space.

## Computation and Bioinformatics

This facility will require computational and bioinformatics capabilities at a variety of levels ranging from sample tracking, automation, instrument control, and the acquisition, analysis, and storage of data to highly sophisticated tools for enabling biological insight. In this group, we

focused mainly on computation and bioinformatics needs associated with isolation and identification of protein complexes, which are in their own right considerable, given the large volumes of data generated by current MS-based proteomics experiments.

Three issues critical to computation must be considered. First is the type of computer hardware to be included in the facility. Second is the infrastructure needs of a facility-wide LIMS. Finally, this facility will require a vibrant computation-research component to address matters such as experimental design to isolate and identify protein complexes by computationally predicting a customized protocol for each complex. Computational methods are possible for choosing the most fruitful complexes to target, and research will be performed in MS data analysis. This would include new algorithms for interpreting mass spectra (e.g., de novo sequencing from MS-MS data) and determining measures of confidence for protein assignments based on MS data. Visualization techniques for gleaning biologically relevant information from heaps of data also could be developed.

The hardware required probably will be in the form of clusters. A commercially available LIMS for a facility of this size would cost millions of dollars. The other possibility is in-house LIMS development, which also would be expensive. A commercial LIMS system will still require in-house information-technology effort. Automation of data handling is an important component of the computing infrastructure.

More generally, a user-friendly Web site or other means of making data output available to the community will be required. This tool must be flexible enough for users to access the output database in the Facility for Whole Proteome Analysis at the level of detail or perspective they want to pursue.

## Summary

Workshop attendees were asked to assess high-level questions that would need to be addressed by this facility. Questions and answers are summarized below.

1. What data are needed by the biological community?
  - Every type of complex should be studied, including protein:protein, protein: DNA, protein:RNA, and protein-small molecule. A complex is defined as any nonrandom association of components in which each contributes to function.
  - Types of data needed include everything-composition and stoichiometry, localization, function, structure, mechanism of action, regulation, assembly, trafficking, information on interaction surfaces and physical chemistry, energetics, and dynamics.
  - Recognizing that "everything" may not be possible, one idea was to have a high-throughput lab generate data identifying potential complexes and have the scientific community do the reductionist science to prove function.
  - Need parallel efforts to characterize transient, low-abundance, low-affinity, and membrane-associated complexes in addition to more stable complexes.
  - Facility products include data on complexes, protocols for isolation and characterization, and training and education
  - Biologists want complex information derived from multiple techniques.
  - An iterative process should allow an ongoing dialog with the community about next steps and what is needed.
  - Experimental determination and computational prediction of complexes will be necessary.
  - The ultimate facility product should be function, not a mere catalog of complexes.
2. How should data and tools be made available to the community?
  - Data should be made available via a Web-based user-friendly interface that can be drilled down for detailed information.
  - Need informatics tools along with data. These tools must be able to integrate molecular information, including DNA, RNA, protein, PTM, and physiological regulation.
  - Community-based data should be made available ASAP.
3. What supporting information is required by the biologist?
  - Computational biologists must develop tools for assigning confidence levels to identified complexes. This information should be linked to the primary data via a Web interface.
  - "Quality-assurance data and process should be available to end users. Documentation of the process must be included as part of the data.
  - "Experimental determination and computational prediction of complexes will be necessary.
  - "Biologists want genetic information for establishing function.
4. What types of samples and requests should be submitted to the facility?
  - Community-driven requests for complete characterization of complexes in a given organism, families of organisms, or genetic variants of certain organisms.
  - Community access to the facility via application to focus on individual pathways or complexes.
  - Different levels of interaction should be considered on an individual basis, driven by the science question being asked and by available methodology.



## Specific Capabilities to be Included in the Protein Complex Facility

- The facility needs to do easy and difficult complexes in parallel.
5. How should the Protein Complex Facility be managed?
- Facility management and advisory board.
  - Training is an important role for the facilities (workshops and protocol manuals).
  - Cater to individuals through proposals.
  - Applications prioritized according to integration with the big picture.
  - Interactions with other GTL facilities are strongly encouraged; this will enable "big" science.
  - "Calls" that seek to integrate facilities should be prioritized.
  - A research presence as well as a high-throughput factory should be part of the facility.
6. How should the Protein Complex Facility interact with other facilities?
- The Protein Production Facility would produce reagents that could be used in the Protein Complex Facility. For example, affinity-tagged proteins grown in *E. coli* could be used as "bait" to isolate complexes from other organisms. Antibodies to complexes and clones expressing fusion proteins could come from the Protein Production Facility.
  - The Proteomics Facility would produce proteomics data to be used by the Protein Complex Facility. Knowing which proteins are expressed by a particular organism under a specified set of growth conditions would enable the Protein Production Facility to target complexes.
  - The Systems Facility would receive data from and provide input to the Protein Complex Facility for choosing organisms, pathways, and complexes to pursue.
7. How do we interface with other GTL Facilities?
- Strong interaction with other GTL facilities will be necessary.
  - Standards for exchanging information will need to be established.
  - Specific information needed from other facilities will be data relating where and when proteins are expressed and methodologies used to obtain these results.
  - "Custom" samples, labels, and reagents will be needed for in situ generation of biomolecular complexes of appropriate quantity and quality.

Additional information is summarized below with respect to capabilities to be included in the facility, impacts on science and DOE missions, and probability for success.

## Specific Capabilities to be Included in the Protein Complex Facility

1. To discover and define protein complexes in microbes:
- Controlled growth of microbes under carefully documented conditions.
  - Affinity reagents (either antibodies or tagged proteins), cross-linking reagents to stabilize complexes, and others that will be employed to isolate protein complexes from microbes.
  - Automated processes, including parallel sample processing, combinatorial affinity "challenges" and others to isolate complexes from cells in a high-throughput manner.
  - Automated, high-throughput identification of isolated complexes using combined chromatographic separations and MS with conventional mass spectrometers and those with high mass-resolving power.
  - Computer algorithms to analyze, assess, archive, and integrate data from multiple measurement techniques.

## Impacts on Science and DOE Missions

2. To analyze the structure and function of molecular machines:
  - Automated imaging tools (cryo-EM, optical and force microscopies, and magnetic resonance imaging) to characterize machines in vitro and establish the relationships of proteins within the complex.
  - High-throughput imaging tools such as optical and magnetic resonance imaging to characterize molecular machines in vivo for establishing the relationships and dynamics of protein complexes within cells.
  - Characterization tools and sensors to generate real-time, multidimensional information for spatial and temporal measurements of protein complexes in vivo.
  - Additional characterization tools such as dynamic and small-angle neutron and X-ray scattering and neutron reflectometry to provide additional biophysical data on isolated complexes.
  - Advanced computational tools and state-of-the-art computer hardware and data storage for analysis and integration of data from various measurements.

## Impacts on Science and DOE Missions

The increased ability to study molecular machines will illuminate the fundamental biochemical and biophysical mechanisms within and on the surfaces of microbial cells. Understanding and harnessing microbial capabilities can revolutionize DOE's methods for accomplishing its energy and environmental-cleanup missions.

- Gaining access to all a cell's machines and to classes of machines across numerous systems will allow in-depth analyses of form, function, and the regulation of key processes.
- Combining computational power with GTL data on tens of thousands of proteins and molec-

ular machines will contribute to solving the important problem of protein refolding.

- Achieving understanding would permit the design in engineered systems of biomimetic chemical processes that match the selectivity, purity, and energy efficiencies of natural reactions.

## Probabilities for Success

With its focus on protein complexes, this facility is a major scientific and technical challenge. Today's technologies admittedly will be applicable to the analysis of many types of protein complexes. However, to meet GTL program needs, great improvements will be necessary to allow the analysis of protein complexes on the same scale as current genome sequencing. Success will depend on R&D and pilots—a mix of existing and new technologies—to realize truly high-throughput assays for the confident identification and characterization of protein complexes.

- New imaging tools and reagents to measure protein machine dynamics and structure. Requires higher spatial resolution, faster data acquisition, improved spectral information (possible without fluorescent tags), and rapid data analysis.
- Assembly or capture of large protein complexes.
- New approaches to stabilize and analyze transient complexes.
- Enhanced capabilities for high-throughput MS analysis with high dynamic range, high sensitivity, and fast data analysis.
- Sophisticated modeling and simulation capabilities for understanding the interactions and processes underlying machine structure and behavior.
- Innovative molecular biology tools for microbial systems.
- Novel approaches for culturing of microbes.
- Automated sample handling to minimize contamination and sample size.
- Highly specific and inexpensive antibody libraries.

# Agenda

## Tuesday, June 17, 2003

- 1:30 p.m. Welcome – Michelle Buchanan, ORNL
- 1:35 p.m. Overview of GTL program – David Thomassen, DOE-OBER
- 1:55 p.m. Protein Complexes – Biological Importance – Frank Larimer, ORNL
- 2:15 p.m. Facility III Overview – Michelle Buchanan, ORNL
- 2:35 p.m. Organism Growth – Bob Kelly, North Carolina State University
- 2:55 p.m. Break
- 3:15 p.m. Complex Isolation, Purification – Robert Tabita, Ohio State University
- 3:35 p.m. Complex Identification
- LCMS – Andrew Link Vanderbilt University School of Medicine Imaging – Bill Margolin, University of Texas Medical School
  - Affinity Methodologies – Joanna Albala, Lawrence Livermore National Laboratory
- 4:35 p.m. Computational/Modeling – Russ Finley, Wayne State University
- 4:55 p.m. Summary and Charge to Breakout Session One – Michelle Buchanan, ORNL
- 5:00 p.m. Breakout Session One – “Defining the Needs of the Biological Community”
- 6:00 p.m. Adjourn
- 6:30 p.m. Social (no host bar)
- 7:30 p.m. Dinner (catered)

## Wednesday, June 18, 2003

- 7:30 a.m. Continental Breakfast
- 8:00 a.m. Continue Breakout Session One
- 9:15 a.m. Reassemble to hear reports from each Breakout Group on “Defining the Needs of the Biological Community” (3 groups, 15 min. each)
- 10:00 a.m. Charge for Breakout Session Two, “Technology Needs” – Michelle Buchanan, ORNL
- 10:15 a.m. Breakout Sessions
- Growth and Expression of Protein Complexes
  - High Throughput Isolation and Identification
  - Imaging/Biophysical Characterization/Validation
- 12:00 p.m. Lunch (catered)
- 1:00 p.m. Breakout Session Two, “Technology Needs,” continues
- 2:30 p.m. Reassemble to hear reports from Breakout Session Two, “Technology Needs” (20 minutes each)
- Growth and Expression of Protein Complexes
  - High Throughput Isolation and Identification
  - Biophysical Characterization/Validation
- 3:30 p.m. Cross-Cut – Computing – Ed Uberbacher, ORNL
- 3:50 p.m. Summary/Wrap up
- 4:00 p.m. Adjourn

June 11, 2003

To: Confirmed Participants of the ORNL GTL Facility III Workshop:

Thank you for participating in ORNL's GTL Facility III Workshop next Tuesday and Wednesday, June 17 and 18. The meeting will start promptly at 1:30 p.m. in the Sinclair Amphitheater in the Atlanta Airport Hilton. The goals for the workshop are to:

- Assess the needs of the biological community that will drive Facility III
- Define initial capabilities needed
- Define technology needs of the Facility
- Plan operational aspects of the Facility

In addition to plenary talks, two breakout sessions will be held. The first breakout will address **Defining the Needs of the Biological Community** and will include consideration of the following issues.

1. What key biological questions could be answered with new information from Facility III on the comprehensive identification and characterization of protein complexes?
2. What type of data or knowledge does the biological community need that should be provided by Facility III? [This is a general question that should identify technologies at a high level that will be discussed in separation breakout sessions.]
3. What information, supporting data, is needed for the biologist to have confidence in the data?
4. How should the data be made available to the biological community? What types of computational tools need to be available?
5. In what form(s) should samples/requests be submitted to the Facility? Should the Facility establish standard experimental and analytical "templates" or protocols to guide users on the capabilities of the Facility?
6. What skills, background, and capabilities will biologists need to make use of comprehensive data that Facility III will provide?

The second breakout will focus on technology needs for Facility III and will be subdivided into three working groups: **Growth and Expression of Protein Complexes**, **High Throughput Isolation and Identification**, and **Biophysical Characterization/Validation**. Issues to be addressed include:

1. What are the technologies needed for Facility III? [This is in the context of each specific breakout session.]

2. What science breakthroughs are required to enable the goals of the facility? What R&D is needed to fill gaps in current technologies that will be needed prior to Facility III assuming operation? What are longer term (5-10+ years) technology needs?
3. How can these technologies be scaled up for a high throughput production mode?
4. What are computational/bioinformatics needs associated with these technologies?
5. What quality assurance/validation/standardization procedures are needed in conjunction with technologies, data sets, sample preparation, and other aspects to ensure that the Facility provides the biological community highly valid information?

Cross-cutting issues, such as computing and QA needs, will be assessed throughout the workshop. A draft Agenda for the workshop is attached. Please note on the agenda that we have arranged for catered meals for invited participants. If you have any special dietary requests, please let Brenda Campbell know by COB Thursday, June 12, so that we may accommodate your request.

To facilitate our discussions, I ask that you familiarize yourself with the DOE mission challenges and the how the Underlying goals of the Genomes to Life program will support those mission challenges. You can access this information on the DOE GTL web site at <http://doegenomestolife.org/>. Additional information you may find useful includes the attached materials:

- "Genomes to Life: Realizing the Potential of the Genome Revolution," a six-page overview of GTL and the facilities.
- "User Facilities for 21st Century Systems Biology: Providing Critical Technologies for the Research Community." A PDF file of which the back 30 pages describes the facilities in the most currently available detail.
- "Facility Interactions," A PowerPoint file that depicts how Facility III interacts with other GTL facilities.

For your convenience, we have posted these materials on our web site at <http://www.ornl.gov/GenomestoLife/news/2003meeting.html>.

We look forward to your participation and to a productive workshop. Please feel free to contact me or Brenda Campbell (campbellbw@ornl.gov, 865/574-4860) if you have any questions.

Sincerely,



Michelle Buchanan, Director  
Chemical Sciences Division

## Summary

Workshop attendees were asked to assess high-level questions that would need to be addressed by this facility. Questions and answers are summarized below.

1. What data are needed by the biological community?
  - Every type of complex should be studied, including protein:protein, protein: DNA, protein:RNA, and protein-small molecule. A complex is defined as any nonrandom association of components in which each contributes to function.
  - Types of data needed include everything-composition and stoichiometry, localization, function, structure, mechanism of action, regulation, assembly, trafficking, information on interaction surfaces and physical chemistry, energetics, and dynamics.
  - Recognizing that "everything" may not be possible, one idea was to have a high-throughput lab generate data identifying potential complexes and have the scientific community do the reductionist science to prove function.
  - Need parallel efforts to characterize transient, low-abundance, low-affinity, and membrane-associated complexes in addition to more stable complexes.
  - Facility products include data on complexes, protocols for isolation and characterization, and training and education
  - Biologists want complex information derived from multiple techniques.
  - An iterative process should allow an ongoing dialog with the community about next steps and what is needed.
  - Experimental determination and computational prediction of complexes will be necessary.
  - The ultimate facility product should be function, not a mere catalog of complexes.
2. How should data and tools be made available to the community?
  - Data should be made available via a Web-based user-friendly interface that can be drilled down for detailed information.
  - Need informatics tools along with data. These tools must be able to integrate molecular information, including DNA, RNA, protein, PTM, and physiological regulation.
  - Community-based data should be made available ASAP.
3. What supporting information is required by the biologist?
  - Computational biologists must develop tools for assigning confidence levels to identified complexes. This information should be linked to the primary data via a Web interface.
  - Quality-assurance data and process should be available to end users. Documentation of the process must be included as part of the data.
  - Experimental determination and computational prediction of complexes will be necessary.
  - Biologists want genetic information for establishing function.
4. What types of samples and requests should be submitted to the facility?
  - Community-driven requests for complete characterization of complexes in a given organism, families of organisms, or genetic variants of certain organisms.
  - Community access to the facility via application to focus on individual pathways or complexes.
  - Different levels of interaction should be considered on an individual basis, driven by the science question being asked and by available methodology.

**ORNL GTL Facility III Workshop Attendees**

**June 17-18, 2003  
Atlanta, Georgia**

<b>Last Name</b>	<b>First Name</b>	<b>Organization</b>	<b>E-mail</b>	<b>Phone</b>
Finley	Russell	Wayne State Univ	<a href="mailto:rfinley@wayne.edu">rfinley@wayne.edu</a>	313-577-7845
Foote	Bob	ORNL	<a href="mailto:footers@ornl.gov">footers@ornl.gov</a>	865-576-2032
Fornes	Raymond	NCSU	<a href="mailto:refornes@unity.ncsu.edu">refornes@unity.ncsu.edu</a>	919-515-7865
Frank	Ed	ANL	<a href="mailto:efrank@mcs.anl.gov">efrank@mcs.anl.gov</a>	630-252-4548
Giddings	Michael	UNC	<a href="mailto:mgidding@email.unc.edu">mgidding@email.unc.edu</a>	919-843-3513
Gorin	Andrei	ORNL	<a href="mailto:agor@ornl.gov">agor@ornl.gov</a>	865-241-3972
Gunter	M. Jean	Ga Tech	<a href="mailto:jean.gunter@oars.gatech.edu">jean.gunter@oars.gatech.edu</a>	404-894-6908
Hansis	Nancy	Flad & Associates	<a href="mailto:nancy_hansis@flad.com">nancy_hansis@flad.com</a>	608-232-4307
Harris	Frank	ORNL	<a href="mailto:harrisf@ornl.gov">harrisf@ornl.gov</a>	865-574-4333
Hazen	Terry	LBNL	<a href="mailto:tchazen@lbl.gov">tchazen@lbl.gov</a>	510-486-6223
Horton	Linda	ORNL	<a href="mailto:hortonll@ornl.gov">hortonll@ornl.gov</a>	865-574-5081
Hu	Xiaoping	Ga Tech	<a href="mailto:xhu@emory.edu">xhu@emory.edu</a>	404-712-2615
Hurst	Greg	ORNL	<a href="mailto:hurstgb@ornl.gov">hurstgb@ornl.gov</a>	865-574-6142
Kelly	Bob	NCSU	<a href="mailto:rmkelly@eos.ncsu.edu">rmkelly@eos.ncsu.edu</a>	919-515-6396
Kennel	Steve	ORNL	<a href="mailto:kennelsj@ornl.gov">kennelsj@ornl.gov</a>	865-574-0825
Kirby	John	Ga Tech	<a href="mailto:john.kirby@biology.gatech.edu">john.kirby@biology.gatech.edu</a>	404-894-8418
Knotek	Mike	Consultant	<a href="mailto:m.knotek@verizon.net">m.knotek@verizon.net</a>	520-877-3133
Lane	Steve	LLNL	<a href="mailto:lane17@llnl.gov">lane17@llnl.gov</a>	925/423-7463

**ORNL GTL Facility III Workshop Attendees**

**June 17-18, 2003  
Atlanta, Georgia**

<b>Last Name</b>	<b>First Name</b>	<b>Organization</b>	<b>E-mail</b>	<b>Phone</b>
Larimer	Frank	ORNL	<a href="mailto:larimerfw@ornl.gov">larimerfw@ornl.gov</a>	865-574-1253
Link	Andrew	Vanderbilt	<a href="mailto:andrew.link@vanderbilt.edu">andrew.link@vanderbilt.edu</a>	615-343-6823
Liphardt	Jan	LBNL	<a href="mailto:jliphardt@lbl.gov">jliphardt@lbl.gov</a>	510-642-1440
Liu	Jun	SNL	<a href="mailto:jliu@sandia.gov">jliu@sandia.gov</a>	505-845-9135
Masingo	Susan	ORNL	<a href="mailto:masginosw@ornl.gov">masginosw@ornl.gov</a>	865-574-4333
Mansfield	Betty	ORNL	<a href="mailto:mansfieldbk@ornl.gov">mansfieldbk@ornl.gov</a>	865-576-6669
Margolin	William	U of Texas @ Houston	<a href="mailto:william.margolin@uth.tmc.edu">william.margolin@uth.tmc.edu</a>	713-500-5452
McGlashan	Dave	Perkins & Will	<a href="mailto:dave.mcglashan@perkinswill.com">dave.mcglashan@perkinswill.com</a>	404-443-7521
Michaels	George	PNNL	<a href="mailto:george.michaels@pnl.gov">george.michaels@pnl.gov</a>	
Mukhopadhyay	Biswarup	Va Tech	<a href="mailto:biswarup@vt.edu">biswarup@vt.edu</a>	540-231-8015 540-231-1219
Nagamori	Gary	HDR	<a href="mailto:gnagamor@hdrinc.com">gnagamor@hdrinc.com</a>	650-943-2025
Nie	Shuming	Ga Tech & Emory	<a href="mailto:snie@emory.edu">snie@emory.edu</a>	
Palzkill	Timothy	Baylor College of Medicine	<a href="mailto:timothyp@bcm.tmc.edu">timothyp@bcm.tmc.edu</a>	713-798-5009
Pelletier	Dale	ORNL	<a href="mailto:pelletierda@ornl.gov">pelletierda@ornl.gov</a>	865-576-2857
Peterson	Cynthia	UT-K	<a href="mailto:cynthia_peterson@utk.edu">cynthia_peterson@utk.edu</a>	865-974-4083
Resnekov	Orna	The Molecular Sciences Inst.	<a href="mailto:resnekov@molsci.org">resnekov@molsci.org</a>	510-981-8747
Rintoul	Danny	SNL	<a href="mailto:rintoul@sandia.gov">rintoul@sandia.gov</a>	505-844-9592
Rodland	Karin	PNNL	<a href="mailto:karin.rodland@pnl.gov">karin.rodland@pnl.gov</a>	509-376-7605



**ORNL GTL Facility III Workshop Attendees**

**June 17-18, 2003  
Atlanta, Georgia**

<b>Last Name</b>	<b>First Name</b>	<b>Organization</b>	<b>E-mail</b>	<b>Phone</b>
Schachat	Frederick	Duke	<a href="mailto:f.schachat@cellbio.duke.edu">f.schachat@cellbio.duke.edu</a>	919-684-4106
Scott	Robert	UGA	<a href="mailto:rscott@uga.edu">rscott@uga.edu</a>	706-542-2726
Singh	Anup	SNL	<a href="mailto:aksingh@sandia.gov">aksingh@sandia.gov</a>	925-294-1260
Spicer	Len	Duke	<a href="mailto:spicer@biochem.duke.edu">spicer@biochem.duke.edu</a>	919-684-4327
Strausberg	Robert	TIGR	<a href="mailto:rls@tigr.org">rls@tigr.org</a>	301-838-5890
Tabita	Robert (Bob)	OSU	<a href="mailto:tabita.1@osu.edu">tabita.1@osu.edu</a>	614-292-4297
Thomassen	David	DOE-OBBER	<a href="mailto:david.thomassen@science.doe.gov">david.thomassen@science.doe.gov</a>	301-903-9817
Trehwella	Jill	LANL	<a href="mailto:jtrehwella@lanl.gov">jtrehwella@lanl.gov</a>	505-667-2690
Uberbacher	Ed	ORNL	<a href="mailto:uberbacherec@ornl.gov">uberbacherec@ornl.gov</a>	865-574-6134
Verberkmoes	Nathan	UT-K/GST	<a href="mailto:verberkmoesn@ornl.gov">verberkmoesn@ornl.gov</a>	865-574-2848
Wang	Bi-cheng (B.-C)	UGA	<a href="mailto:wang@bc11.bmb.uga.edu">wang@bc11.bmb.uga.edu</a>	706-542-1747
Wang	Yisong	ORNL	<a href="mailto:wangy1@ornl.gov">wangy1@ornl.gov</a>	865-574-5396
Wolf	Matthew	Ga Tech	<a href="mailto:matthew.wolf@cc.ga.tech.edu">matthew.wolf@cc.ga.tech.edu</a>	404 385-1278
Xu	Dong	ORNL	<a href="mailto:xud@ornl.gov">xud@ornl.gov</a>	865-574-8934
Xu	Ying	ORNL	<a href="mailto:xuy1@ornl.gov">xuy1@ornl.gov</a>	865-574-7263
Zhou	Jizhong (Joe)	ORNL	<a href="mailto:zhouj@ornl.gov">zhouj@ornl.gov</a>	865-576-7544