Structural Proteomics Workshop
Natcher Building, Rooms G1 and G2
National Institutes of Health
Bethesda, Maryland

MONDAY, APRIL 7, 2003

8:00 - 8:30 a.m.	Registration
8:30 - 8:45 a.m.	Wah Chiu, Workshop Coordinator Workshop Goals
8:45 - 9:00 a.m.	Judy Vaitukaitis, NCRR Director Welcome Remarks

Session I: Cell Biology and Complexes

9:00 - 9:45 a.m. Discussion leader: Paul Matsudaira

David Drubin
Why Studying Large Protein Complexes is Crucial at this Time

9:45 – 10:15 a.m. Break

Session II: Biological Processes and Protein Machines

10:20 a.m. - 12:00 noon Discussion leader: David Drubin

Jack Greenblatt
Protein Complexes in Transcription

Peter K. Sorger
Chromosome Segregation and Kinetochore Structure in Yeast

Rebecca Heald
Protein Complexes in Cell Division in Eggs from Xenopus Laevis

12:00 noon - 1:00 p.m. Lunch Break (Boston Room, Natcher Building)
Box lunches provided for all participants

Session III: Genetic and Chemical Approaches

1:00 - 2:30 p.m. Discussion leader: Trisha Davis

John Yates
Analysis of Protein Complexes by Mass Spectrometry
Charlie Boone
Synthetic Genetic Array Analysis: Large-Scale Mapping of Genetic Interactions Networks in Yeast
William Jacobs
Mining Relevance from the Mycobacterium Tuberculosis Genome and Structural Database

2:30 – 3:00 p.m. Break

Session IV: Computational Approaches

3:00 - 4:15 p.m. Discussion leader: Helen Berman

Mark Gerstein
Computational Proteomics: Prediction of Protein Interactions and Function on a Genome-Scale

Andrej Sali
Modeling the Structures of Molecular Assemblies by Satisfaction of Spatial Restraints

Session V: Electron Cryomicroscopy

4:15 - 5:30 p.m. Discussion leader: Wah Chiu

Ken Downing
Electron Crystallography: Current State and Prospects

Ron Milligan
The Combined Use of Cryo-Electron Microscopy and X-Ray Crystallography to Understand the Structure and Mechanisms of Molecular Machines

7:15 p.m. Dinner (Local restaurant to be determined)

TUESDAY, APRIL 8, 2003

Session VI: Crystallography

8:30 - 9:45 a.m. Discussion leader: David Davies

Steve Almo
Structural Proteomics: Structural Dissection of Protein Machines

David Stuart
Structural Proteomics of Infectious Agents

Session VII: Proteomics Centers

9:45 - 10:45 a.m. Discussion leader: Lee Makowski

Alex Kurosky
Proteomics in Airway Inflammation

Michelle Buchanan
New Approaches for High-Throughput Identification and Characterization of Protein Complexes

Trisha Davis
Yeast Resource Center

10:45 – 11:15 a.m. Break

Session VIII: Recommendations

11:15 a.m. - 1:00 p.m. Discussion leader: Steve Almo

Lee Makowski
Realistic Goals and Milestones; Resources Needed and Funding Mechanisms

1:00 Adjourn

Structural Proteomics Workshop
April 7-8, 2003
Participant List

Steve Almo
Biochemistry
Albert Einstein College of Medicine
(718) 430-2746
almo@aecom.yu.edu

Helen Berman
Chemistry and Chemical Biology
Rutgers University
(732) 445-4667
berman@rcsb.rutgers.edu

Charlie Boone
Department of Medical Research
University of Toronto
(416) 946-7260
charlie.boone@utoronto.ca

Michelle Buchanan
Chemical Sciences Division
Oak Ridge National Laboratory
(865) 574-4986
buchananmv@ornl.gov

Jean Chin
Division of Cell Biology & Biophysics
NIGMS, NIH
(301) 594-2485
chinj@nigms.nih.gov

Wah Chiu
Biochemistry and Molecular Biology
Baylor College of Medicine
(713) 798-6985
wah@bcm.tmc.edu

Dean Cole
Medical Sciences Division, SC-73
U.S. Dept. of Energy
(301) 903-3268
dean.cole@science.doe.gov

David Davies
Laboratory of Molecular Biology
NIDDK, NIH
(301) 496-4295
david.davies@nih.gov

Trisha Davis
Biochemistry
University of Washington
(206) 543-5345
tdavis@u.washington.edu

Jim Deatherage
Cell Biology and Biophysics
NIGMS, NIH
(301) 594-3828
deatherj@nigms.nih.gov

Kenneth Downing
Life Sciences Department
Lawrence Berkeley National Lab
(510) 486-5941
khdowning@lbl.gov

David Drubin
Molecular and Cell Biology
U.C. Berkeley
(510) 642-3692
drubin@uclink4.berkeley.edu

Charles G. Edmonds

Cell Biology and Biophysics
NIGMS, NIH
(301) 594-4428
edmondsc@nigms.nih.gov

Greg Farber
Biomedical Technology
NCRR, NIH
(301) 435-0755
farberg@ncrr.nih.gov

Paula Flicker
NIGMS, NIH
(301) 594-3828
flickerp@nigms.nih.gov

Mark Gerstein
Mol. Biochemistry & Biophysics
Yale University
(203) 432-6105
Mark.Gerstein@yale.edu

Jack Greenblatt
Dept. of Medical Research
University of Toronto
(416) 978-4141
jack.greenblatt@utoronto.ca

Rebecca Heald
Molecular and Cell Biology
University of California, Berkeley
(510) 643-5493
heald@socrates.berkeley.edu

Sonia Hess
LBC
NIDDK, NIH
(301) 402-2197
Sonja_Hess@nih.gov

Roland F. Hirsch
Medical Sciences Division, SC-73
U.S. Department of Energy-GTN
(301) 903-9009
roland.hirsch@science.doe.gov

William R. Jacobs, Jr.
Microbiology & Immunology
Howard Hughes Medicine Institute at
Albert Einstein College of Medicine
(718) 430-2888
jacobsw@hhmi.org

Warren Jones
PPBC Division
NIGMS, NIH
(301) 594-5938
jonesw@nigms.nih.gov

Karl Koehler
Molecular and Cellular Biosciences
National Science Foundation
(703) 292-8440
kkoehler@nsf.gov

Boyana Konforti
Senior Editor, Cell & Mol. Cell;
Executive Editor, Structure
Cell Press
(617) 397 2825
bkonforti@cell.com

Alexander Kurosky
Human Biological Chemistry &
Genetics
Univ. of Texas Medical Branch
Galveston
(409) 772-2771
akurosky@utmb.edu

Catherine Lewis
CBB, Biophysics
NIGMS, NIH
(301) 594-0828
lewisc@nigms.nih.gov

Jerry Li
Cell Biology and Biophysics Division
NIGMS, NIH
(301) 451-4317

lij@nigms.nih.gov

Lee Makowski
Biosciences Division
Argonne National Laboratory
(630) 252-3819
lmakowski@anl.gov

Paul Matsudaira
Biological Engineering
Whitehead Institute, MIT
(617) 258-5188
matsudaira@wi.mit.edu

Michael Marron
Biomedical Technology
NCRR, NIH
(301) 435-0755
marronm@ncrr.nih.gov

Ron Milligan
Cell Biology
The Scripps Research Institute
(858) 784-9827
milligan@scripps.edu

John Norvell
NIGMS, NIH
(301) 594-0533
norvellj@nigms.nih.gov

Bret Peterson
Biomedical Technology
NCRR, NIH
(301) 435-0758
bretp@ncrr.nih.gov

Louise Ramm
NCRR, NIH
(301) 435-0879
RammL@mail.nih.gov

David Robinson
NHLBI, NIH
(301) 435-0477
dr14j@nih.gov

Andrej Sali
Biopharmaceutical Sciences and
Pharmaceutical Chemistry
and California Institute for
Quantitative Biomedical Research
UCSF
(415) 514 4227
salil@salilab.org

Douglas Sheeley
Division of Biomedical Technology
NCRR, NIH
(301) 594-9762
ds8q@nih.gov

Peter Sorger
Biology and Biological Engineering
Massachusetts Inst. of Technology
(617) 252-1648
psorger@mit.edu

David Stuart
Division of Structural Biology
University of Oxford
+44 1865 287546
Dave@strubi.ox.ac.uk

Amy Swain
NCRR, NIH
(301) 435-0752
swaina@ncrr.nih.gov

David Thomassen
Biological & Environmental Research
U.S. Department of Energy
(301) 903-9817
david.thomassen@science.doe.gov

Marjorie Tingle
Biomedical Technology
NCRR, NIH
(301) 435-0772
MarjorieT@ncrr.nih.gov

Valda Vinson

Editor, Science Magazine
(202) 326-6571
vvinson@aaas.org

Judith Vaitukaitis
Director
NCRR, NIH
(301) 496-5793
JudyV@NCRR.NIH.GOV

John Yates
Cell Biology
The Scripps Research Institute
(858) 784-8862
jyates@scripps.edu

Jin Kim Yeoun
NIDDK, NIH
(301) 496-1200
yeounjin_kim@nih.gov

Why Studying Large Protein Complexes is Crucial at this Time

DAVID G. DRUBIN

Large protein complexes are important for virtually every biological process and most disease states. Complete genome sequences have revealed 10,000s of genes in a variety of different organisms. The daunting task of understanding the functions and regulation of these genes can be simplified by associating each of these genes with one of 200-300 core biological processes. Since most proteins function in association with other proteins in cells, one way to simplify the problem of assigning functions to the many genes identified in genome sequences is to identify their biologically relevant interactions, and to elucidate the fundamental properties of the complexes within which proteins function. A host of new and/or improved technologies and experimental approaches makes this an appropriate time for such studies. These approaches include, but are not limited to, mass spectrometry, two-hybrid screening, systematic genetic analysis, crystallographic and cryo-EM structure analysis, expression profiling, and a variety of light microscopy techniques made especially valuable by the ability to express GFP fusions in live cells. Combining approaches provides a means to validate results and to extract biological meaning. It is critical in all such endeavors to remember that the focus needs to be on increasing our understanding of biological mechanisms and the basis for disease.

Protein Complexes in Transcription

JACK GREENBLATT

We are using tandem affinity purification (TAP) tags (protein A + calmodulin binding peptide) to tag and purify each of the ~6200 proteins of *S. cerevisiae,* and have carried out purifications from more than 2000 tagged strains thus far. We are also using somewhat similar SPA tags (3 X FLAG + CBP) to tag and purify each of the ~4000 proteins of *E. coli.* The success rate thus far for more than 600 *E. coli* proteins, of which about one-third are conserved, essential, and potential targets for new broad spectrum antibiotics, has been 76%.

After two steps of affinity chromatography, purified proteins are identified by SDS-PAGE followed by protein band excision and MALDI-ToF mass spectrometry, as well as by direct trypsin digestion of the eluate from the last column followed by LC/MS/MS analysis. Once approximately stoichiometric components of the various purified complexes are identified, a clustering algorithm is used to organize the bait/prey data into a series of independent protein complexes. Once the order of the prey proteins is rearranged so that the protein complex clusters fall along the diagonal of a 2D plot, sub-stoichiometric interacting polypeptides are added to the data, enabling us to directly visualize interactions between protein complexes. In principle, therefore, it will be possible to use the information in our database to decide which protein complexes to mix together in order to determine co-structures for interacting protein complexes.

We have especially focused on protein complexes involved in transcription by RNA polymerase II and in the production and processing of stable RNAs and mRNAs. In the latter case, many distinct complexes have been identified with distinct roles in various processing events in the production of ribosomal RNA. In the former case, we began by purifying all the proteins thought to have a role in elongation by RNA polymerase II. This procedure enabled us to identify many new polypeptides with a role in elongation (e.g. the Paf1 complex, consisting of 5 polypeptides) and to establish the beginnings of an interaction network that connects them. Our purification of the Set1 and Set2proteins, which methylate histone H3, established that the Set1 complex (COMPASS; 8 polypeptides) interacts with the Paf1 complex and the Set2 protein with RNA polymerase II, thereby extending the elongation factor network. We used synthetic genetic arrays (SGA) to identify still other genes with putative roles in elongation and tagged and purified those proteins. This further extended our network of polypeptides with putative roles in transcriptional elongation to encompass more than 100 polypeptides. Systematic chromatin immunoprecipitation (ChIP) is being used to localize the various protein complexes to particular regions along transcribed genes.

In order to organize the protein complexes into functional groups by a different criterion and to assess whether the polypeptides of a given complex have similar or distinct biological roles, we have again turned to SGA analysis with the intention of examining the phenotypes of the

~20,000 possible deletion double mutants in our network. Preliminary clustering of the genetic interaction data has revealed that functional "genetic groups," which presumably represent biochemical pathways, are sometimes the same as and sometimes distinct from protein complexes.

Chromosome Segregation and Kinetochore Structure in Yeast

PETER SORGER

Kinetochores are multi-protein complexes that assemble on centromeric DNA and mediate the attachment and movement of chromosomes along the microtubules (MTs) of the mitotic spindle. I will discuss recent work from our laboratory on the simplest eukaryotic centromeres and kinetochores, those found in the budding yeast *S. cerevisiae*. Research on kinetochore function and chromosome segregation is focused on four questions of general significance: what specifies the location of centromeres, what are the protein components of kinetochores and how do they assemble a MT attachment site, how do MT attachments generate force, and how do cells sense the state of attachment via the spindle assembly checkpoint?

In the last few years, significant progress has been made in identifying the protein components of yeast kinetochores and in determining their overall architecture. Until quite recently, yeast kinetochores were thought to comprise a small set of DNA binding proteins and a kinesin related protein (KRP). We now know that budding yeast kinetochores contain at least 50 subunits, including multiple MT binding proteins. About half of these proteins are encoded by essential genes and several have very close mammalian homologues. We are currently attempting to build a complete list of kinetochore components, encompassing perhaps 60-80 proteins, and determine the roles these proteins in chromosome-microtubule attachment to be determined in cells. Biochemical and in vivo studies are underway to elucidate protein-protein interactions, assembly intermediates and the approximate positions of various proteins in the DNA-to-MT bridge. We believe that these studies set the stage for detailed biophysical analysis of microtubule attachment, force generation and checkpoint signaling.

Protein Complexes in Cell division in Eggs from Xenopus Laevis

REBECCA HEALD

Understanding how an exact copy of the genome is transmitted to daughter cells during cytokinesis is a key question in biology, yet a systematic identification of cell division proteins has been lacking. To obtain an accessible and enriched source of cell division proteins, we isolated mammalian midbodies and identified 159 midbody proteins by tandem liquid chromatography and mass spectrometry (LC/LC/MS/MS). To quickly assess function, we systematically inactivated the homologues of the mammalian proteins in *C. elegans* using RNA-mediated interference (RNAi). Approximately 89.9% of the midbody proteins were conserved between mammals and nematodes and 68% had no previously known function in cell division. Surprisingly, the majority of the midbody proteins function in membrane-cytoskeletal dynamics, and depletion of the individual *C. elegans* homologues revealed a very high percentage of cytokinesis defects (38%) and incomplete germline cytokinesis defects (38%), in addition to an unexpected number of neuronal/uncoordinated defects (16%). The very high degree of conservation and the diverse array of proteins identified point to common and ancient mechanisms mediating cell division, membrane dynamics and neuronal trafficking events, all of which are critical in human development and defective in cancer and/or neurodegenerative diseases.

Analysis of Protein Complexes by Mass Spectrometry

JOHN YATES

The completion of genome projects for the human genome and many model organisms represents a significant step forward for the study of biology. Genome sequences provide the basic blueprint for the functional potential of an organism and provide an initial description of the genes and gene products. Proteins often work together in complexes to perform essential functions such as protein translation, signal transduction, gene transcription, gene splicing and other major functions of the cell. Deciphering the components of protein complexes can help us to understand how these molecular machines perform these functions and how they are regulated. Mass spectrometry has become a key tool to determine the components of complexes. Initial strategies have involved the use of electrophoretic techniques in combination with mass spectrometry to identify the proteins. Newer strategies have involved the direct analysis of the digested complexes to identify the components and more recently to identify modifications to the proteins. Mass spectrometers are evolving at a fast rate to become faster and more sensitive with the ability to measure m/z values with higher resolution and better mass accuracy. All these improvements will lead to more precise measurements of complex composition, modifications and stoichiometry.

Computational Proteomics:
Prediction of Protein Interactions and Function on a Genome-Scale

MARK GERSTEIN

My talk will address a major post-genomic challenge: trying to predict protein function on a genomic scale. I will approach both of this through analyzing the properties and attributes of proteins in a database framework. The work on predicting protein function will discuss the strengths and limitations of a number of approaches: (i) using sequence similarity; (ii) using structural similarity; (iii) clustering microarray experiments; and (iv) data integration. The last approach involves systematically combining information from the other three and holds the most promise for the future. For the sequence analysis, I will present a similarity threshold above which functional annotation can be transferred, and for the microarray analysis, I will present a new method of clustering expression timecourses that finds "time-shifted" relationships.

References

J. Qian, B. Stenger, C. Wilson, J. Lin, R. Jansen, W. Krebs, V. Alexandrov, N. Echols, S. Teichmann, J. Park, and M. Gerstein. "PartsList: a web-based system for dynamically ranking protein folds based on disparate attributes, including whole-genome expression and interaction information." *Nucleic Acids Res.* 29: 1750-64 (2001).

J. Qian, M. Dolled-Filhart, J. Lin, H. Yu, and M. Gerstein. "Beyond synexpression relationships: Local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions." *J. Mol. Biol.* 314: 1053-1066 (2001).

R. Jansen, D. Greenbaum, and M. Gerstein. "Relating whole-genome expression data with protein-protein interactions." *Genome Res.* 12: 37-46 (2002).

http://bioinfo.mbb.yale.edu/e-print/integ-complex/all.pdf
http://bioinfo.mbb.yale.edu/e-print/expint-genomeres/reprint.pdf
http://bioinfo.mbb.yale.edu/e-print/timeshift-jmb/reprint.pdf
http://bioinfo.mbb.yale.edu/e-print/tongperspec/reprint.pdf
http://bioinfo.mbb.yale.edu/e-print/integ-datamine-ieee/proofs.pdf

Modeling the Structures of Molecular Assemblies by Satisfaction of Spatial Restraints

ANDREJ SALI

There is a need for a computational approach to derive the best three-dimensional model of an assembly, at the appropriate resolution, given all of the available experimental data and theoretical considerations. Such modeling can be seen as an optimization problem whose solution requires satisfaction of spatial restraints. The approach depends on a representation of the modeled system, an objective function consisting of spatial restraints, and an optimizer that produces a model by minimizing violations of the input restraints. We implemented modeling by satisfaction of spatial restraints in the MODELLER program. Although the program has been applied primarily to comparative modeling, it also provides a convenient tool in the modeling of molecular assemblies. In addition to describing applications to two specific assemblies (below), several considerations for a combined experimental and theoretical mapping of the universe of macromolecular complexes will be discussed.

Comparative protein structure models for molecular modeling of the yeast ribosome. While a combination of cryo-electron microscopy (cryoEM) and X-ray crystallography allows us to determine atomic structures of macromolecular assemblies, this process is time consuming. Fortunately, the structures of large assemblies and their constituent parts tend to be conserved in evolution, enabling homology or comparative protein structure modeling to enhance the value of low-resolution electron density maps. Comparative modeling can currently be used to characterize the 3D structure at the level of a fold for approximately 30% of all domains in the known sequences. This fraction will grow rapidly, perhaps by ~10% per year for the next few years, in large part due to structural genomics. Therefore, there is a practical need to improve the use of homologous subunit structures in the modeling of protein assemblies. As an illustration, we present here the fitting of homology-based protein structure models into a cryo-electron density map of the whole yeast and *E. coli* ribosomes.

Modeling the three-dimensional structure of the yeast nuclear pore complex. Some large assemblies, such as the nuclear pore complex (NPC), consist predominantly of subunits whose structures have not yet been defined. Such assemblies may be characterized only by very low-resolution information about their overall shape and some protein-protein contacts. In these cases, we can expect to be able to model only the configuration of the proteins in the assembly, not their individual conformations. We characterized the configuration of the proteins in the yeast NPC (F. Alber, M. Rout, B. Chait, in preparation). The NPC proteins (nucleoporins) are represented as spheres with the radii estimated from their numbers of amino acid residues. The NPC structure was obtained by minimizing violations of the following restraints: exclusion volume restraints, protein-protein proximity restraints extracted from 41 immuno-purification experiments, axial and radial positional restraints on nucleoporins obtained by electron microscopy of gold-coated antibodies

against tagged nucleoporins, restraints imposed by the shape of the nuclear envelope that contains the NPC, and restraints on the symmetry of the NPC derived primarily from electron microscopy at approximately 200Å resolution. Starting with random configurations of the nucleoporins, many 3D models of the NPC were calculated by MODELLER. The final configurations that satisfied the input restraints well were clustered, resulting in one dominant model of the NPC.

Electron Crystallography: Current State and Prospects

KENNETH H. DOWNING

Electron crystallography has developed as a viable technique for determining protein structures at atomic resolution in cases where monolayer crystals of the protein can be obtained. The success of the technique is illustrated with structures of bacteriorhodopsin and tubulin, two proteins that had resisted extensive attempts to obtain crystals suitable for x-ray crystallography. Electron diffraction methods have also been applied with these two proteins to understand much more than just the protein structure. In the case of bacteriorhodopsin, conformational changes associated with several stages in the photocycle have been elucidated using Fourier difference methods applied to samples trapped in these states. With tubulin, we have been studying the binding of microtubule-stabilizing drugs that could have anti-cancer potential similar to Taxol.

As of this date, though, only a few structures have been determined by electron crystallography. At this point, x-ray crystallography has advantages over electron crystallography that make it the preferred choice when suitable crystals can be obtained. Data collection and structure solution are both much more rapid, and the resolution obtained is generally higher. Advances in instrumentation and methodology for electron crystallography should soon narrow the gap between the two techniques, making it much more efficient to obtain structural information on proteins with which high quality, two-dimensional crystals can be obtained.

The Combined Use of Cryo-Electron Microscopy and X-Ray Crystallography to Understand the Structure and Mechanisms of Molecular Machines

RON MILLIGAN

Although the ongoing efforts in structural genomics are producing detailed pictures of individual proteins and nucleic acids – the immediate products of the genome - it is apparent that many cellular functions are accomplished by multicomponent complexes or molecular machines. Individual protein and nucleic acid molecules may be regarded as the "nuts and bolts" from which molecular machines are built. Understanding the structure and mechanisms of action of these machines represents the next frontier in efforts to define how the gene products function in a cellular context.

Many molecular machines are present in the cell in low copy number, and they are likely to be compositionally and conformationally dynamic. They are therefore extremely challenging objects for study by x-ray crystallographic or NMR methods alone. However, high resolution crystallographic or NMR studies on individual components or stable subcomplexes, together with lower resolution cryo-electron microscopy on the entire complex, offers a means to build near atomic models of the machines and to understand their molecular mechanisms.

To illustrate both the general principles of the approach and the sorts of structural and functional information that it may yield, I will present some results of cryo-EM studies on kinesin motors. Low-resolution 3D maps of kinesin-microtubule complexes provided a framework into which the atomic resolution structures of the separate components were docked. Relative orientations of the components in the complex were accurately defined by a gold-cluster labeling / difference mapping strategy. Trapping of the complex at various stages in its enzymatic cycle, together with specific gold cluster labeling revealed the conformational changes in the motor that result in directional motion along the underlying microtubule. From these data we can put together a detailed picture of a machine at work.

Although these studies were done on a relatively simple molecular machine, the approach is applicable to many of the cellular complexes that have recently been identified. I will conclude with a discussion of what I view as the challenges we face in applying this approach to a wide variety of macromolecular machines.

Proteomics in Airway Inflammation

ALEX KUROSKY

The University of Texas Medical Branch at Galveston (UTMB) is one of ten Proteomic Centers in the U.S. established by the National Heart, Lung, and Blood Institute (NHLBI) to enhance and develop innovative proteomic technologies and apply them to relevant biological questions relating to heart, lung, blood and sleep health and disease. The NHLBI Proteomic Initiative is intended to complement and enhance ongoing research programs currently supported by the NHLBI. The ten Proteomic Centers will be supported through a contract mechanism over seven years for a total of $157 million. Each Center proposes to develop and validate proteomic technologies targeted to specific biological questions. Early review of the activities of the Centers clearly underscored the need to develop better strategies for identifying and analyzing protein complexes in comparative studies, e.g. control versus disease or drug treated. Several technologies are currently employed to characterize protein complexes, as for example, tandem liquid chromatography in conjunction with tandem mass spectrometry of immunoprecipitates, confocal microscopy for co-localization studies, as well as protein chip methods. However, there is an obvious need to further develop these new instrumentations to advance their capabilities and sensitivities and to make them accessible to researcher in order to facilitate protein complex studies. NCRR support for these instruments for applications to proteome analysis is strongly encouraged.

One example of a study of protein complex characterization ongoing at UTMB in Dr. Allan Brasier's laboratory relates to the regulation of the transcription factor nuclear factor-□B (NF-□B) by the I□B kinase (IKK) complex. This multi-protein complex consists of the kinases IKK-? and IKK-?, the regulatory IKK-? subunit, as well as other unidentified protein components and kinases. The need to characterize and identify these other unknown proteins is pivotal to a full understanding of NF-□B regulation. The characterization of the IKK complex will initially focus on A549 cells infected with Respiratory Syncytial Virus (RSV). RSV is a mucosal-restricted virus that is the leading cause of epidemic respiratory tract infections in children worldwide for which no vaccine is available. RSV replication activates IKK, producing inflammation in the airway and wheezing in infected children. A major goal of this project is to comparatively characterize the IKK complex in viral vs control A549 cells to identify specific inducible or viral-related proteins in the IKK complex. These proteins may subsequently become useful targets for drug inhibitor development to control the inflammatory response to viral infection. Thus, it is clearly an opportune time to introduce the structural characterization of protein complexes into the infrastructure of institutional research. The NCRR support mechanism could significantly impact on the enhancement of protein complex structural studies.

New Approaches for High-Throughput Identification and Characterization of Protein Complexes

MICHELLE V. BUCHANAN

Proteins seldom act in isolation, but combine to form multi-protein complexes that function as "molecular machine" that are responsible for virtually all activities of a cell. Therefore the identification and characterization of these complexes are a key step in understanding cellular function. Our current knowledge of protein complexes is quite limited because proteins are conventionally studied individually, in isolation. Identifying and characterizing of all the molecular machines of a cell represent a substantial technical challenge, however, because these complexes are highly dynamic, changing in amount, modification state, stoichiometry and subcellular location to carry out the primary functions of a cell. In a pilot project sponsored by the DOE Office of Biological and Environmental Research as part of their Genomes to Life program, we have assembled a multidisciplinary team of biologists, chemists, and computational scientists to develop new approaches for the high throughput identification and characterization of protein complexes in microbial cells. This includes evaluation of new techniques for isolating ("pulling down") complexes from cells and analysis by mass spectrometry, imaging and other analytical tools. Informatics tools are being integrated "cradle to grave" as cells are cultured, samples are analyzed, and data is processed and archived in databases that will be made available to the scientific community. Computational tools are also being developed to assist in data interpretation, prediction and modeling. Initial studies are being conducted using two model organisms, *R. palustris* and *S. Oneidensis,* to evaluate the tools being developed and incorporated into a high throughput analysis process.

Yeast Resource Center

http://depts.washington.edu/~yeastrc

TRISHA DAVIS

The Yeast Resource Center is a Biomedical Technology Resource Center supported by the NCRR. The mission of the center is to exploit the genome sequence to facilitate the identification and characterization of protein complexes in the yeast *Saccharomyces cerevisiae*. We provide expertise and access to four advanced technologies: mass spectrometry, two-hybrid arrays, fluorescence microscopy including fluorescence resonance energy transfer, and protein structure prediction. Through collaborations with YRC personnel these critical, but costly, technologies are made available to a large community of researchers throughout the country. Currently, we have approximately 80 collaborative projects. The investigators participating in the YRC include Dr. John Yates and Dr. Ruedi Aebersold for mass spectrometry, Dr. Stan Fields for two-hybrid analysis, Dr. Trisha Davis and Dr. Eric Muller for fluorescence microscopy, and Dr. David Baker for protein structure prediction.

A major challenge for the future is the identification of the function of the many unknown proteins predicted by genome sequences. Two contrasting tactics are a high-throughput approach, which can sacrifice quality for quantity, and the individual analysis of each protein, which can progress slowly. In the YRC, we are using a third approach that seeks to strike a balance between the quality and the quantity of the results. We have applied all four technologies to characterize the proteins encoded by the essential unknown open reading frames in yeast. These 100 proteins are not homologous to any proteins of known function, although many share regions of similarity with unknown proteins in other organisms. The project is almost complete, and several conclusions are clear. First, optimization of the protocols for the purification of the unknown proteins greatly improved the ability to discriminate between contaminants and specifically interacting proteins based on mass spectrometric data. Second, the integration of the different technologies, each with its own strength, increases the confidence in the functional predictions and broadens their significance.