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

Center for Molecular and Cellular Systems

Michelle V. Buchanan
H. Steven Wiley, Frank W. Larimer
Oak Ridge National Laboratory
Pacific Northwest National Laboratory

Collaborating Laboratories
Argonne National Laboratory, Sandia National Laboratories,
University of Utah,
University of North Carolina
Team Leaders
Core:
  Steven Kennel, Thomas Squire
High Throughput Complex Processing
  Mike Ramsey, Karin Rodland
Mass Spectrometry
  Greg Hurst, Richard Smith
Molecular and Cellular Imaging
  Mitch Doktycz, Steve Colson
Bioinformatics and Computing
  Ying Xu, David Dixon

Carol Giometti (ANL) gel electrophoresis
Ray Gesteland (U. Utah) mass spectrometry
Malin Young (SNL) cross-linking
Mike Giddings (U. North Carolina) mass spectrometry
Goal 1
“Identify and Characterize the Molecular Machines of Life”

“...instead of a cell dominated by randomly colliding individual protein molecules, we now know that nearly every major process in a cell is carried out by assemblies ... of proteins...Indeed an entire cell can be viewed as a factory that contains an elaborate network of interlocking assembly lines, each of which is composed of a set of large protein machines.”


Protein complexes are key to biological function

Understand the network of reactions that occur in sufficient detail to predict, test, and comprehend the responses of a biological system to changes

Center for Molecular & Cellular Systems
Goal 1 includes three main steps

- Identify complement of protein complexes and their components
- Elucidate function and dynamics of complexes—intermediates, nature of interactions, cellular location, kinetics
- Establish how changes arising from environmental stress, development, etc., affect complex formation and function

which lay the foundation for GTL
Impact of Goal 1

- Molecular level understanding of protein complexes and, ultimately, networks
- Predict/change behavior of organism and community
- Predict function, biological pathways by homology
- Discover new functions
Identification and Characterization of Protein Machines

- New approaches needed for large-scale studies
  - No single tool will provide all required information
  - Computational tools must be integrated from beginning
    - Analyze, compare, predict, share data
    - Quality assessment
    - Guide experimental design and data collection

Develop integrated approach to correlate identified complexes with data from gene expression, protein expression, imaging, and other methods

Center for Molecular & Cellular Systems
Strategy to Achieve Goal 1

- Initiate protein complex identification using affinity separation combined with mass spectrometry and computational tools
  - Use multiple approaches, non-optimized techniques
  - Focus on targeted complexes

- Evaluate new approaches for high-throughput identification
  - Identify bottlenecks, opportunities for automation
  - Establish dynamic R&D program to develop new, integrated analytical and computational tools

- Incorporate additional tools, data to characterize complexes
  - Imaging tools to characterize complexes in cells
  - Tools to identify interaction interfaces
An Approach for High Throughput Identification of Protein Complexes

Combine complex isolation, mass spectrometry and data analysis

- Bioinformatics
- Cloning, tagging
- Controlled cell growth
- Affinity isolation
- scFv production
- Cross-linking
- Separation
- MS analysis
- Data analysis, archival

Center for Molecular & Cellular Systems
Modified pDEST Vectors for Protein Expression in *R. palustris*

- **pBBR-N’GST-15**
  - RBS ATG GST attR1 CmR ccdB attR2 T7term
- **pBBR-N’HIS-17**
  - RBS ATG 6xHis attR1 CmR ccdB attR2 T7term
- **pBBR-C’GST-24**
  - attR1 CmR ccdB attR2 GST T7term
- **pBBR-C’HIS-42**
  - attR1 CmR ccdB attR2 V5 epitope 6xHis T7term

- **pBBR1MCS5** (Broad host range plasmid)
- **pDEST gateway expression plasmid**

**Center for Molecular & Cellular Systems**
Modified Gateway system for production of affinity tagged \textit{R. palustris} proteins

\begin{align*}
\text{PCR} & \quad \text{pDONOR-221} & \quad \text{BP Clonase} & \quad \text{Entry clone} \\
\text{att} \quad \text{gene} \quad \text{att} & + & \text{att} \quad \text{ccdB} \quad \text{att} & \rightarrow & \text{att} \quad \text{gene} \quad \text{att} \\
\text{Entry clone} & \quad \text{pBBR-DEST} & \quad \text{LR Clonase} & \quad \text{Expression clone} \\
\text{att} \quad \text{gene} \quad \text{att} & + & \text{att} \quad \text{ccdB} \quad \text{att} & \rightarrow & \text{att} \quad \text{gene} \quad \text{att} \\
\end{align*}

\textbf{Protein stain}

\begin{tabular}{|c|c|}
\hline
Expression clone & Vector only \\
\hline

gene

\hline

\end{tabular}

\textbf{Western blot}

\begin{tabular}{|c|c|}
\hline
Expression clone & Vector only \\
\hline

gene

\hline

\end{tabular}

\textbf{Mass Spectrometry}

\begin{tabular}{|l|c|}
\hline
\textbf{GST-nitrite reductase band} & \\
\hline
\textbf{GST peptides} & 10 \\
\textbf{GST/linker peptides} & 1 \\
\textbf{nitrite reductase peptides} & 0 \\
\textbf{linker/nit. red. peptides} & 0 \\
\hline
\textbf{MALDI} & \textbf{22} \\
\hline
\textbf{MS/MS} & \textbf{22} \\
\hline
\end{tabular}
Verification of *R. palustris* Fusion Proteins Expressed in *E. coli*—Two Approaches

- Affinity capture of tagged proteins from lysed cells
  - 1D PAGE
  - Whole eluate digestion
- In-gel digestion and mass spectrometric identification of individual gel bands
- LC-MS-MS of digest peptides; identification of proteins via SEQUEST

<table>
<thead>
<tr>
<th>Fusion Protein</th>
<th>No. of peptides identified from:</th>
<th>Others identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rpal 4709 + N-terminal GST</td>
<td>45, target protein: 8, affinity tag: 2</td>
<td></td>
</tr>
<tr>
<td>Rpal 4709 + C-terminal 6-His &amp; V5 epitope</td>
<td>31, target protein: 3, affinity tag: 19</td>
<td></td>
</tr>
<tr>
<td>Rpal 5426 + C-terminal 6-His &amp; V5 epitope</td>
<td>35, target protein: 3, affinity tag: 8</td>
<td></td>
</tr>
</tbody>
</table>

*These are candidate methods for analysis of protein complexes isolated via affinity purification*
ORNL GTL Process Flowchart

Choose Gene and Growth Conditions

Engineer Tagged Protein

Transfected Cells

Grow Cells Under Specific Conditions

Fractionate Cells

Mass Spec Analysis

- Bottom-Up Analysis
  - Peptide Spectra
- Top-Down Analysis
  - Whole Protein Spectra

Data Analysis

Center for Molecular & Cellular Systems
Analysis of expression of affinity tagged *R. palustris* genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Affinity Tag</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>nirK</td>
<td>Nitrite reductase</td>
<td>N-His</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-His</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-GST</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-GST</td>
<td>0</td>
</tr>
<tr>
<td>groEL-2</td>
<td>chaperonin</td>
<td>N-His</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-His</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-GST</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-GST</td>
<td>+</td>
</tr>
<tr>
<td>groEL-1</td>
<td>chaperonin</td>
<td>N-His</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-His</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-GST</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-GST</td>
<td>+</td>
</tr>
<tr>
<td>soxB</td>
<td>thiosulfate oxidation</td>
<td>N-His</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-His</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-GST</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-GST</td>
<td>0</td>
</tr>
<tr>
<td>soxC</td>
<td>thiosulfate oxidation</td>
<td>N-His</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-His</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-GST</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-GST</td>
<td>+</td>
</tr>
<tr>
<td>hupS</td>
<td>uptake hydrogenase small subunit</td>
<td>N-His</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-His</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-GST</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-GST</td>
<td>++</td>
</tr>
<tr>
<td>hupL</td>
<td>uptake hydrogenase large subunit</td>
<td>N-His</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-His</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-GST</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-GST</td>
<td>++</td>
</tr>
</tbody>
</table>

+++ Excellent  ++ Good  + Poor  0 None
Heterologous Expression

Select Gene

Clone gene

Express & Purify

Antigen

Make scFv

Pull down

Analyse (gel)

Antigen with scFv

Protein complex

PRISM analysis

Center for Molecular & Cellular Systems
Tagged proteins generated to date for pull-down studies of *S. Oneidensis*

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>hydB</td>
<td>periplasmic Fe hydrogenase small subunit</td>
<td>SO3921</td>
</tr>
<tr>
<td>hydA</td>
<td>periplasmic Fe hydrogenase large subunit</td>
<td>SO3920</td>
</tr>
<tr>
<td>napA</td>
<td>periplasmic nitrate reductase</td>
<td>SO0848</td>
</tr>
<tr>
<td>omcA</td>
<td>decaheme cytochrome C</td>
<td>SO1779</td>
</tr>
<tr>
<td>omcB</td>
<td>decaheme cytochrome C</td>
<td>SO1778</td>
</tr>
<tr>
<td>hoxK</td>
<td>quinone-reactive Ni/Fe hydrogenase small subunit precursor</td>
<td>SO2099</td>
</tr>
<tr>
<td>petA</td>
<td>ubiquinol-cytochrome C reductase iron-sulfur subunit</td>
<td>SO0608</td>
</tr>
<tr>
<td></td>
<td>flavocytochrome C flavin subunit</td>
<td>SO3301</td>
</tr>
<tr>
<td></td>
<td>Gfo/Idh/MocA family oxidoreductase</td>
<td>SO3120</td>
</tr>
<tr>
<td></td>
<td>oxidoreductase molybdopterin-binding</td>
<td>SO0715</td>
</tr>
<tr>
<td>nrfC</td>
<td>formate-dependent nitrite reductase</td>
<td>SO0483</td>
</tr>
<tr>
<td></td>
<td>phosphotyrosine protein phosphatase</td>
<td>SO2208</td>
</tr>
<tr>
<td>ptPA</td>
<td>phosphotyrosine protein phosphatase</td>
<td>SO3124</td>
</tr>
<tr>
<td>pTPB</td>
<td>Tyrosine-specific protein phosphatase</td>
<td>SO4476</td>
</tr>
<tr>
<td>cppP</td>
<td>Spheroplast protein precursor</td>
<td>SO2337</td>
</tr>
<tr>
<td>msrA</td>
<td>methionine sulfoxide reductase (isoform A)</td>
<td>SO2588</td>
</tr>
<tr>
<td>msrB</td>
<td>methionine sulfoxide reductase (isoform B)</td>
<td>SO2337</td>
</tr>
<tr>
<td>lniB</td>
<td>ATP-dependent RNA helicase</td>
<td>SO0407</td>
</tr>
<tr>
<td>rpoD</td>
<td>RNA polymerase sigma-70 factor</td>
<td>SO1284</td>
</tr>
<tr>
<td></td>
<td>Cytochrome c3</td>
<td>SO2727</td>
</tr>
<tr>
<td>rpoA</td>
<td>DNA-directed RNA polymerase alpha subunit</td>
<td>SO0256</td>
</tr>
<tr>
<td>rpoZ</td>
<td>DNA-directed RNA omega subunit</td>
<td>SO0360</td>
</tr>
<tr>
<td>hepA</td>
<td>RNA polymerase-associated protein</td>
<td>SO0575</td>
</tr>
</tbody>
</table>
Crosslinking and Mass Spectrometry for Protein Complex Analysis
Oak Ridge National Laboratory, Sandia National Laboratories

- Chemical crosslinking has potential for:
  - Stabilizing “fragile” complexes
  - Providing information on distances between particular residues in proteins or complexes
  - Improving throughput for MS analysis of complexes

- Technical issues currently being addressed:
  - Low abundance of crosslinked products
  - Interpretation of mass spectrometry data
Protein Complex Analysis: Proposed Affinity Crosslinker Approach

- Protein crosslinker
- Linker arms
- Biotin

Computational Analysis

LC MS/MS

Avidin affinity binding

Digest fragments

On-line tryptic digest

On-line microdialysis

Center for Molecular & Cellular Systems
AUTOMATION OF PROTEIN PRODUCTION & ANALYSES

A. Macroscale HT Cloning and Sample Preparation

B. Microscale Sample Production for Mass Spec

C. Lab-On-A-Chip

Center for Molecular & Cellular Systems
Emerging approaches for characterizing protein complexes

Molecular and Cellular Imaging Subproject

- Characterize protein complexes in isolation, within cells, and on cell surfaces/interfaces
- Employ multimodality approaches to molecular imaging—optical probes, molecular recognition force microscopy, afm/optical, (optical)\(n\)
- Validate the composition of protein complexes
- Determine the location of specific complexes at cellular and subcellular locations
- Characterize dynamics, binding forces

Center for Molecular & Cellular Systems
Bioinformatics and Computing

- **Short-term goals**
  - Create infrastructure for sample tracking, data collection and analysis
  - Improve tools for predicting and validating members of protein complexes
  - Build tools for interpreting MS data from cross-linked and modified proteins

- **Long-term goals**
  - Predict protein structures involved in forming complexes
  - Predict function of protein complexes
  - Help build global architecture for integrating data necessary for successful systems biology
Computational Tools Support All Aspects of Center

- sample tracking
- work flow monitoring
- library information management
- data processing, storage, management and transmission
- data communication and technical support

---

**Sample Tracking System**

**Library Information Management System**

**Data Storage, Management, Analysis and Transmission**

**Data from Center, Other Labs, etc.**

**Protein Complex Data Depository**

Community Support

---

Center for Molecular & Cellular Systems
Computational Characterization of Protein Complexes

Experimental conditions

mass spec. data

genomic/proteomic databases

Emerging technologies

functional characterization, dynamics simulation, etc.......

structural characterization of protein complexes

protein interaction map construction

protein identification and protein complex identification

Data interpretation and modeling

PDB

structural genomics

Center for Molecular & Cellular Systems
Acknowledgements