# 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

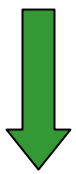
# Goal 1

## *"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."*

***Bruce Alberts,*** *"The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists," Cell*, **92**, 291 (1998)

**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

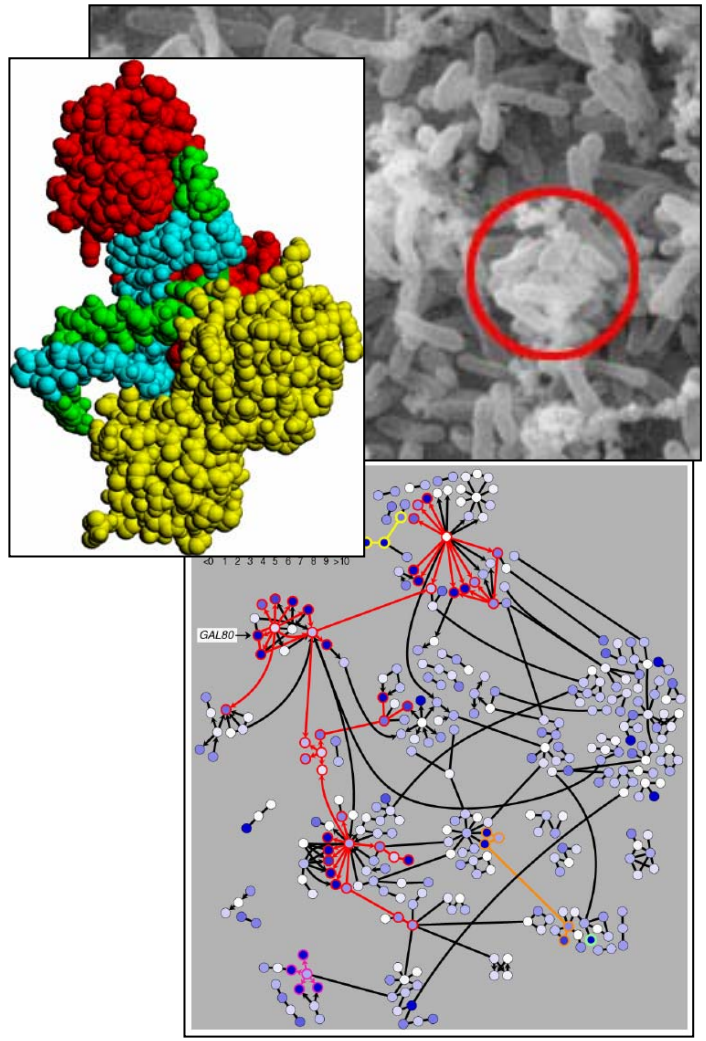
# Goal 1 includes three main steps

- 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

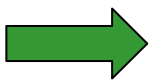
# Impact of Goal 1

- 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

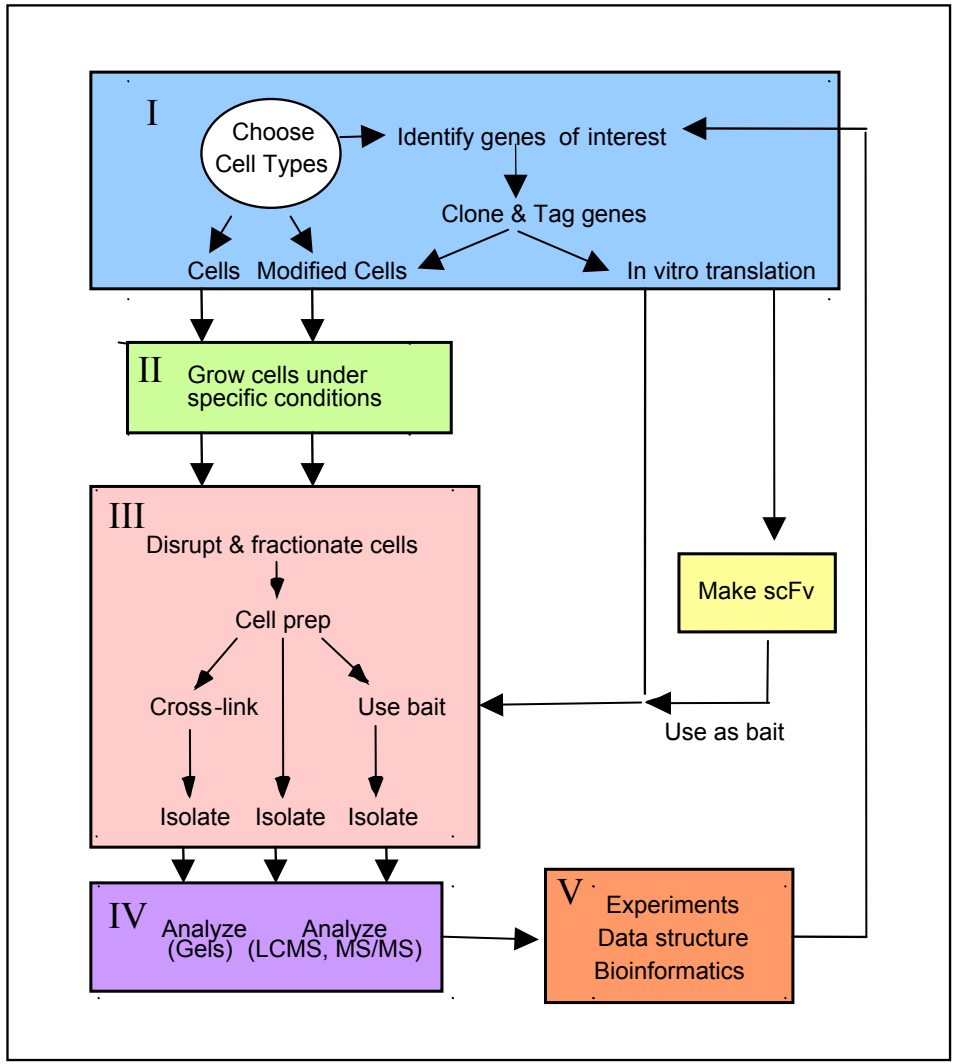
# Strategy to Achieve Goal 1

- 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

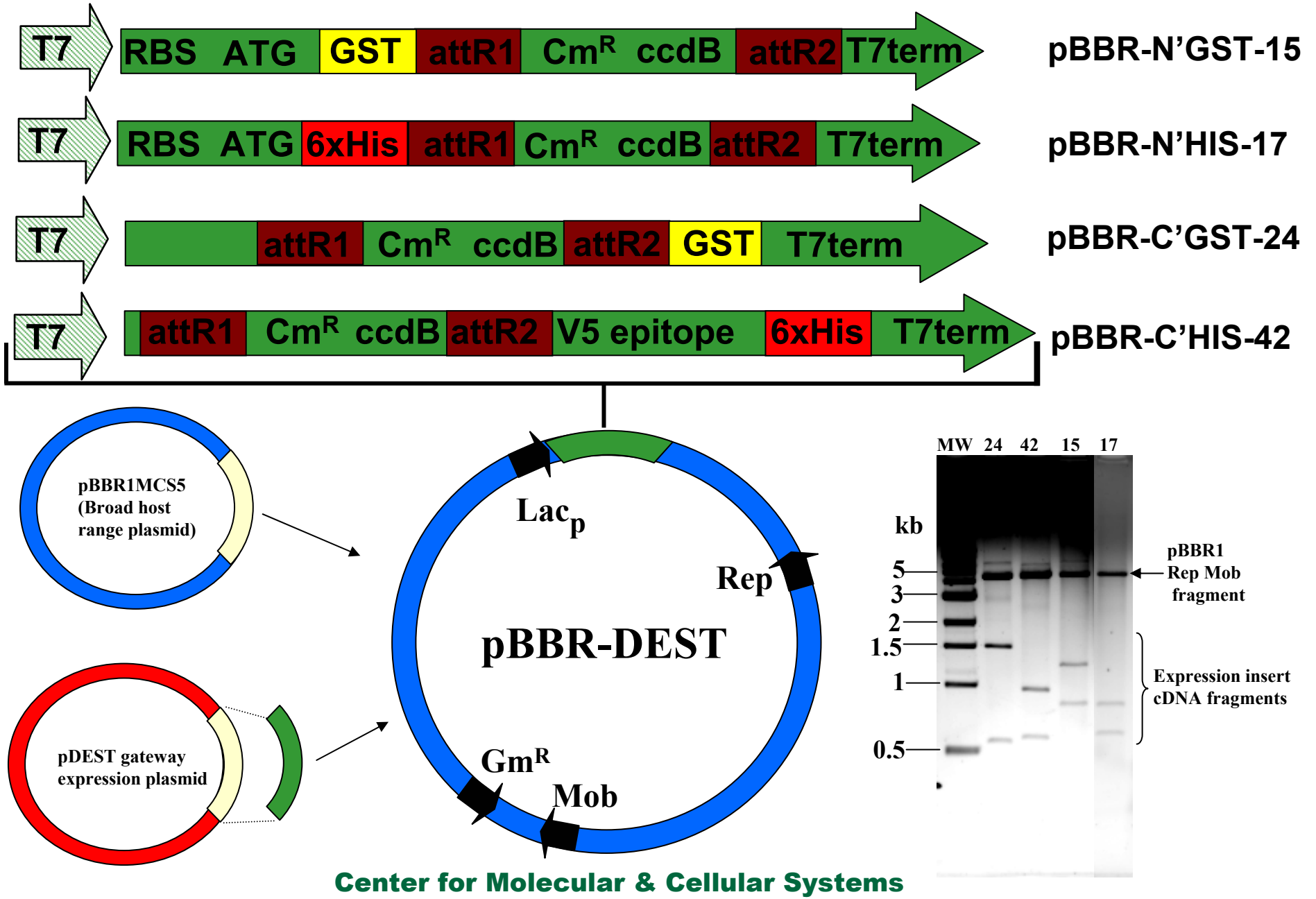
# An Approach for High Throughput Identification of Protein Complexes

**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

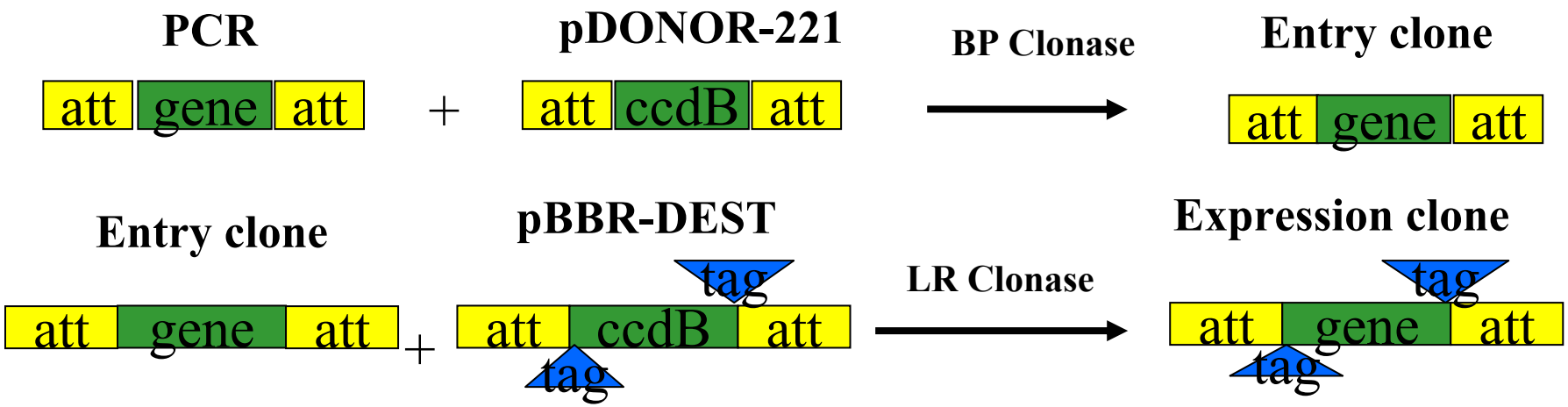


# Modified pDEST Vectors for Protein Expression in *R. palustris*

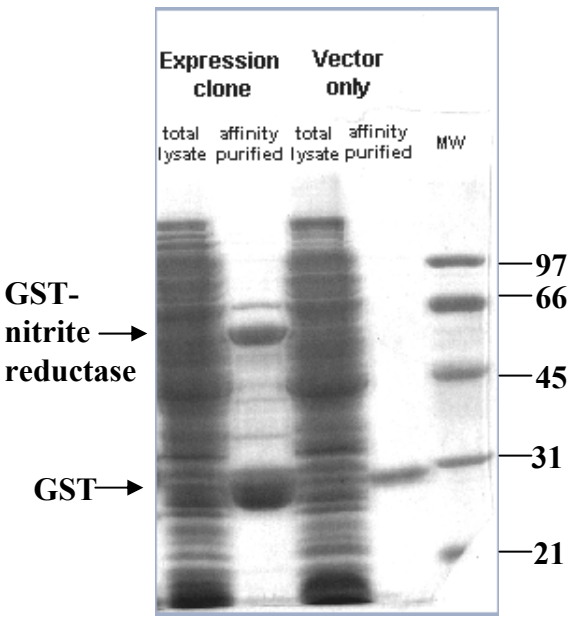



Center for Molecular & Cellular Systems

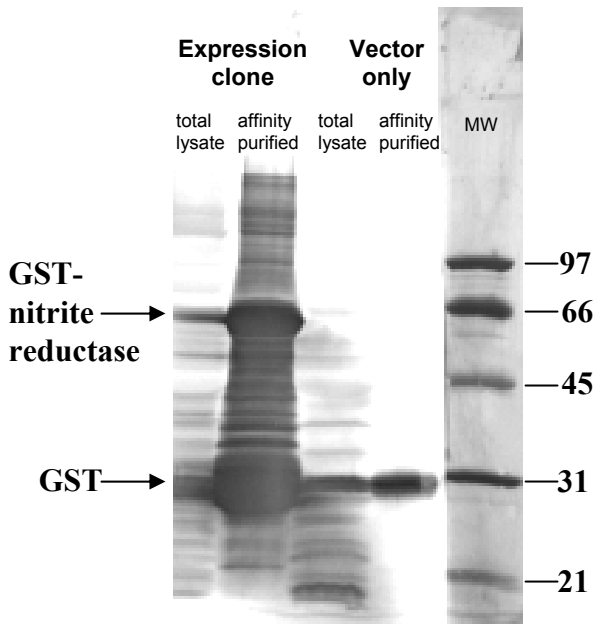
# Modified Gateway system for production of affinity tagged *R. palustris* proteins




## Protein stain




## Western blot




## Mass Spectrometry

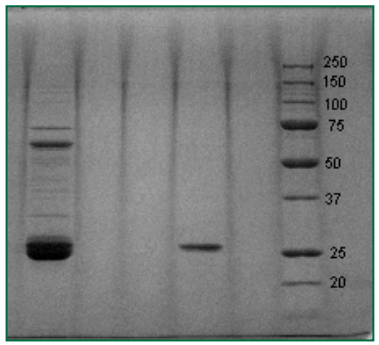
**GST-nitrite reductase band**

| | # of peptides identified | |
|---|---|---|
| | MALDI | nano LC MS/MS |
| GST peptides | 5 | 13 |
| GST/linker peptides | 2 | 1 |
| nitrite reductase peptides | 9 | 11 |
| linker/nit. red. peptides | 0 | 1 |

**GST band**

| | | |
|---|---|---|
| GST peptides | 10 | 22 |
| GST/linker peptides | 1 | 1 |
| nitrite reductase peptides | 0 | 0 |
| linker/nit. red. peptides | 0 | 0 |

# 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

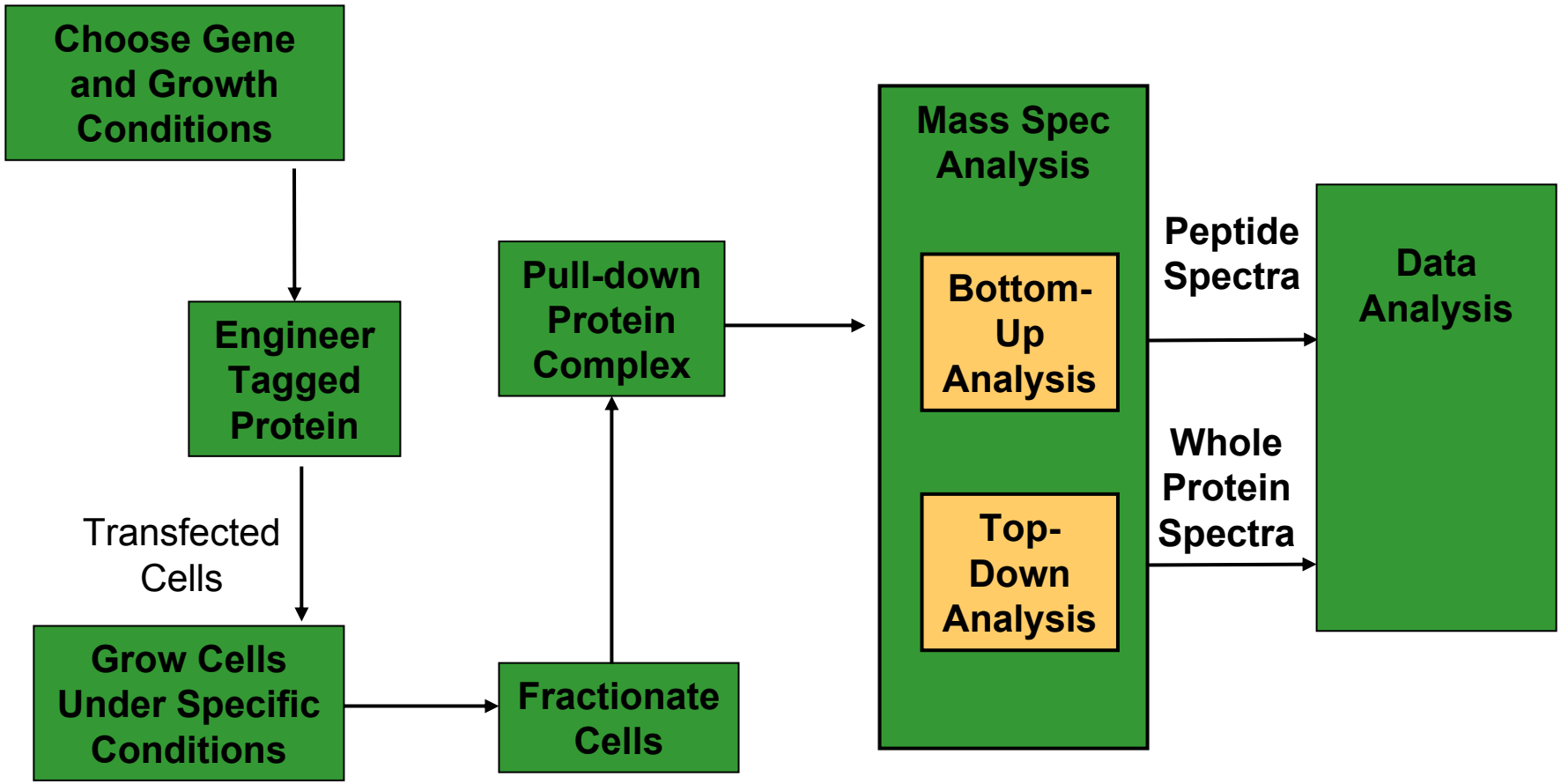
In-gel digestion and mass spectrometric identification of individual gel bands

LC-MS-MS of digest peptides; identification of proteins via SEQUEST

| | No. of peptides identified from: | | Others |
|---|---|---|---|
| **Fusion Protein** | **target protein** | **affinity tag** | ***identified*** |
| Rpal 4709 + N-terminal GST | 45 | 8 | 2 |
| Rpal 4709 + C-terminal 6-His & V5 epitope | 31 | 3 | 19 |
| Rpal 5426 + C-terminal 6-His & V5 epitope | 35 | 3 | 8 |

*These are candidate methods for analysis of **protein complexes** isolated via affinity purification*

# ORNL GTL Process Flowchart

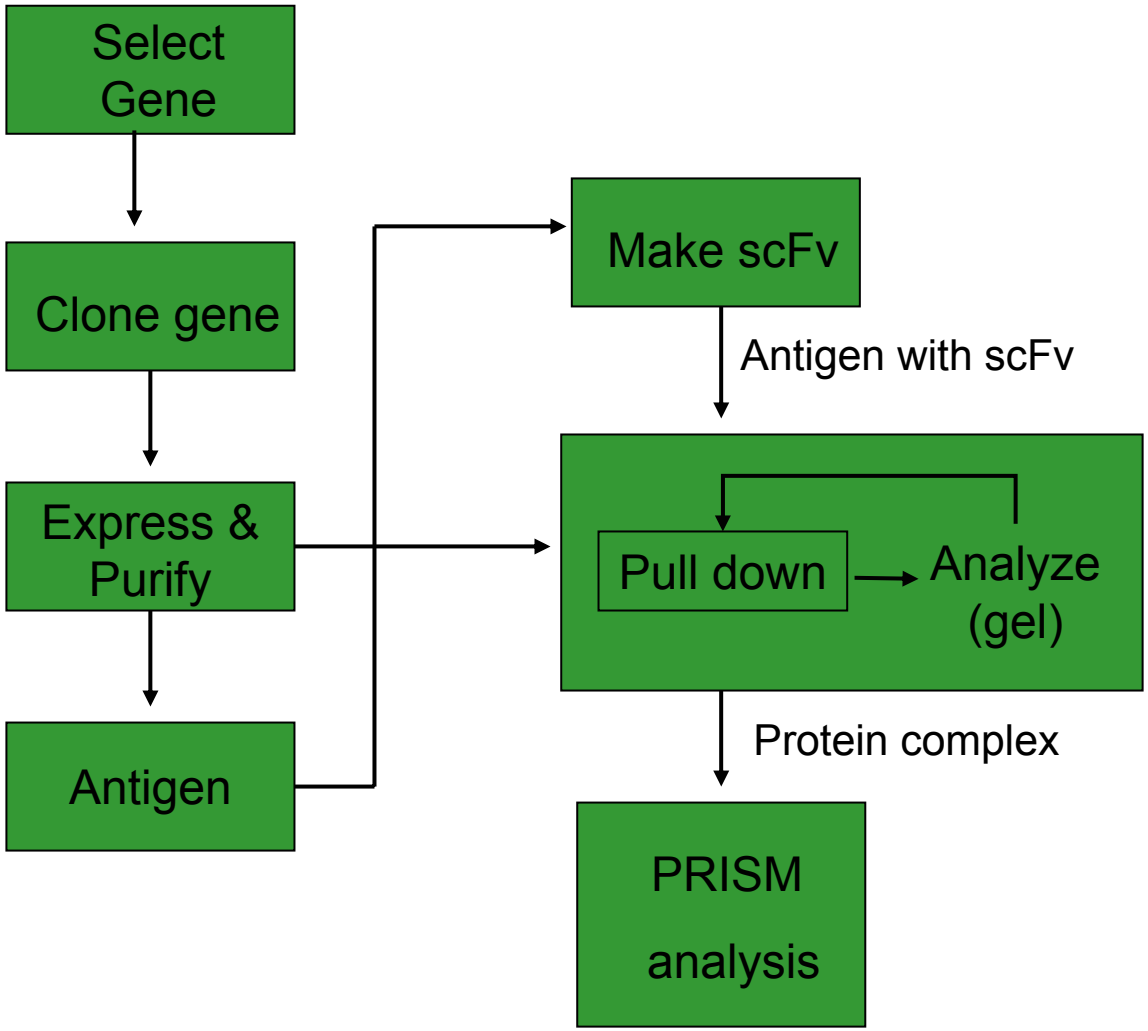


# Analysis of expression of affinity tagged *R. palustris* genes

| Gene | Function | Affinity Tag | Expression |
|---|---|---|---|
| *nirK* | Nitrite reductase | N-His | ++ |
| | | C-His | + |
| | | N-GST | ++ |
| | | C-GST | 0 |
| *groEL-2* | chaperonin | N-His | ++ |
| | | C-His | +++ |
| | | N-GST | + |
| | | C-GST | + |
| *groEL-1* | chaperonin | N-His | +++ |
| | | C-His | +++ |
| | | N-GST | ++ |
| | | C-GST | + |
| *soxB* | thiosulfate oxidation | N-His | ++ |
| | | C-His | ++ |
| | | N-GST | 0 |
| | | C-GST | 0 |
| *soxC* | thiosulfate oxidation | N-His | ++ |
| | | C-His | ++ |
| | | N-GST | + |
| | | C-GST | + |
| *hupS* | uptake hydrogenase small subunit | N-His | 0 |
| | | C-His | ++ |
| | | N-GST | 0 |
| | | C-GST | ++ |
| *hupL* | uptake hydrogenase large subunit | N-His | ++ |
| | | C-His | ++ |
| | | N-GST | 0 |
| | | C-GST | ++ |

+++ Excellent ++ Good + Poor 0 None

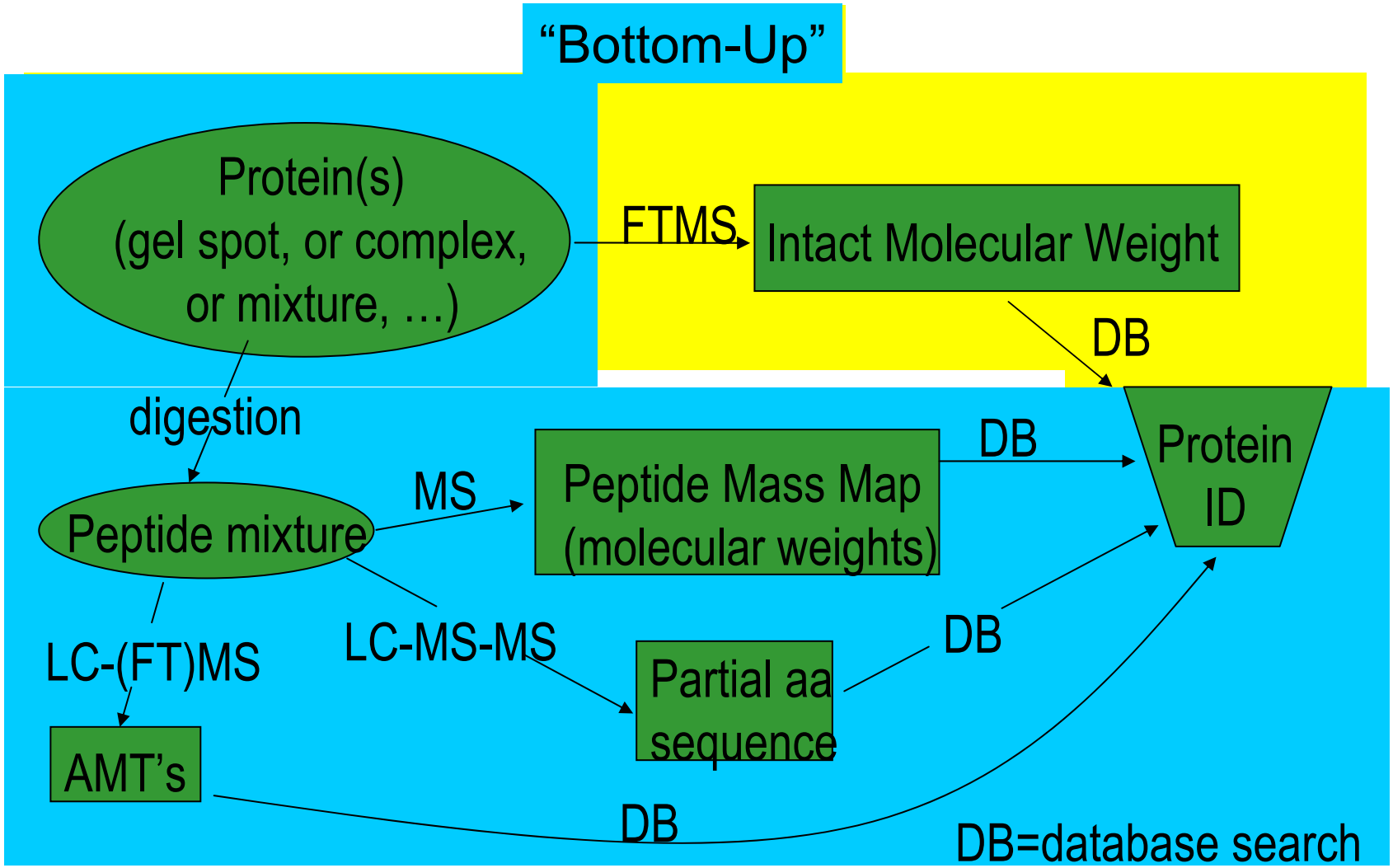
# Heterologous Expression



# Tagged proteins generated to date for pull-down studies of *S. Oneidensis*

| Gene | Description | Annotation |
|---|---|---|
| hydB | periplasmic Fe hydrogenase small subunit | SO3921 |
| hydA | periplasmic Fe hydrogenase large subunit | SO3920 |
| napA | periplasmic nitrate reductase | SO0848 |
| omcA | decaheme cytochrome C | SO1779 |
| omcB | decaheme cytochrome C | SO1778 |
| hoxK | Quinone-reactive Ni/Fe hydrogenase small subunit precursor | SO2099 |
| petA | ubiquinol-cytochrome C reductase iron-sulfur subunit | SO0608 |
| | flavocytochrome C flavin subunit | SO3301 |
| | Gfo/Idh/MocA family oxidoreductase | SO3120 |
| | oxidireductase molybdopterin-binding | SO0715 |
| nrfC | formate-dependent nitrite reductase | SO0483 |
| ptpA | phosphotyrosine protein phosphatase | SO2208 |
| ptpB | Tyrosine-specific protein phosphatase | SO3124 |
| cpxP | Spheroplast protein y precursor | SO4476 |
| msrA | methionine sulfoxide reductase (isoform A) | SO2337 |
| msrB | methionine sulfoxide reductase (isoform B) | SO2588 |
| eno | Enolase | SO3440 |
| rnlB | ATP-dependent RNA helicase | SO0407 |
| rpoD | RNA polymerase sigma-70 factor | SO1284 |
| | Cytochrome c3 | SO2727 |
| rpoA | DNA-directed RNA polymerase alpha subunit | SO0256 |
| rpoZ | DNA-directed RNA omega subunit | SO0360 |
| hepA | RNA polymerase-associated protein | SO0575 |

# MS for Protein Identification



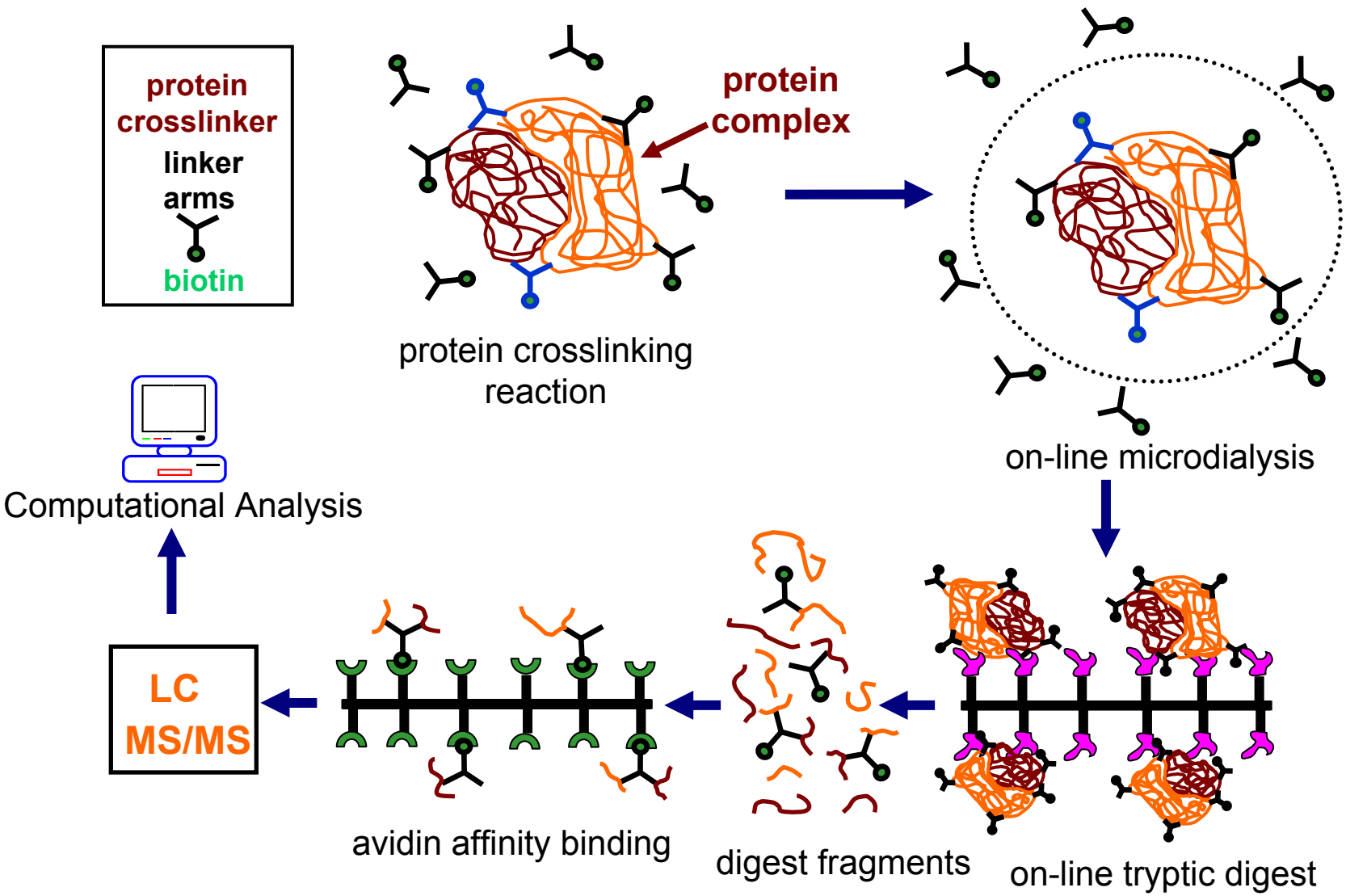

*Integrating "Top-Down" and "Bottom-Up" Mass Spectrometric Approaches for Proteomic Analysis of Shewanella oneidensis*, N.C. VerBerkmoes, J.L. Bundy, L. Hauser, K.G. Asano, J. Razumovskaya, F. Larimer, R.L. Hettich, and J.L. Stephenson, Jr., J. Proteome Research, in press for Vol 1, issue 3 (estimated June 2002).

# Crosslinking and Mass Spectrometry for Protein Complex Analysis

Oak Ridge National Laboratory,
Sandia National Laboratories

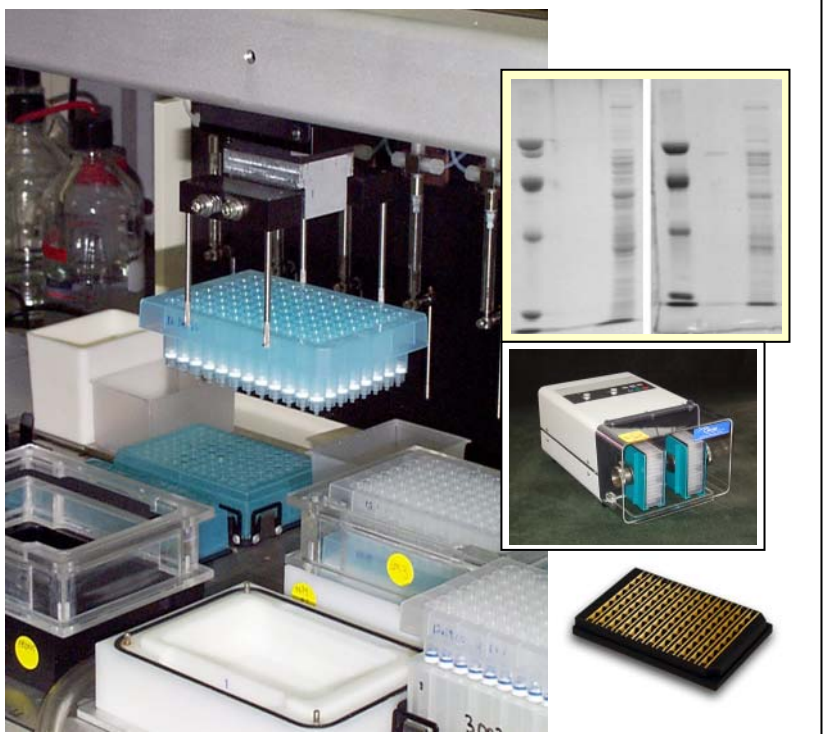
- Chemical crosslinking has potential for:
  - Stabilizing “fragile” complexes
  - Providing information on distances between particular residues in proteins or complexes
  - Improving throughput for MS analysis of complexes
- Technical issues currently being addressed:
  - Low abundance of crosslinked products
  - Interpretation of mass spectrometry data

# Protein Complex Analysis: Proposed Affinity Crosslinker Approach

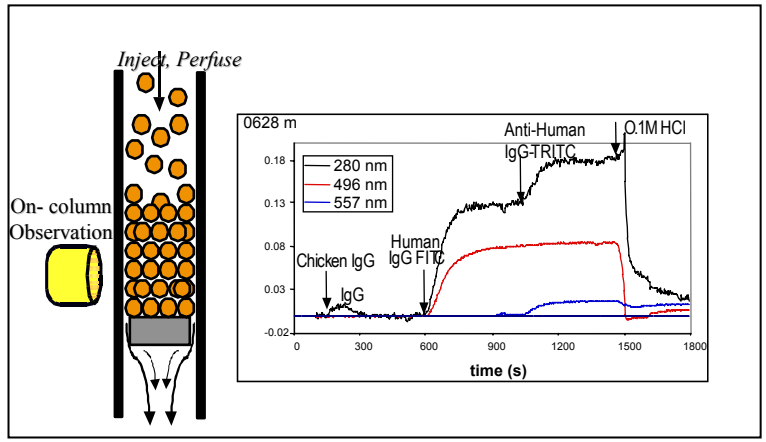


# AUTOMATION OF PROTEIN PRODUCTION & ANALYSES

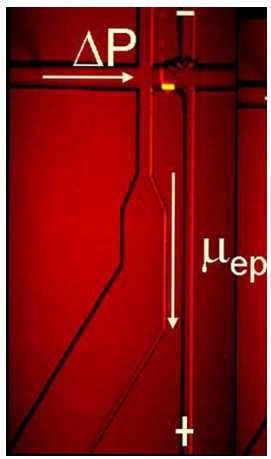
A



B




C



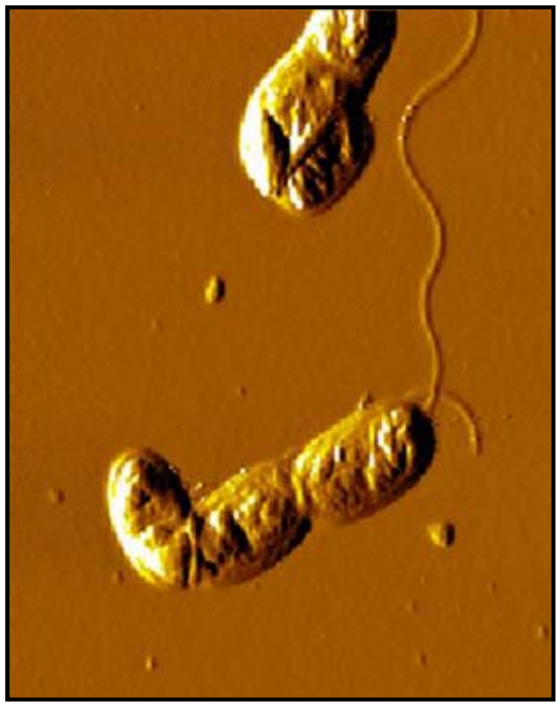

A. **Macroscale HT Cloning and Sample Preparation**

B. **Microscale Sample Production for Mass Spec**

C. **Lab-On-A-Chip**

# Emerging approaches for characterizing protein complexes

## *Molecular and Cellular Imaging Subproject*



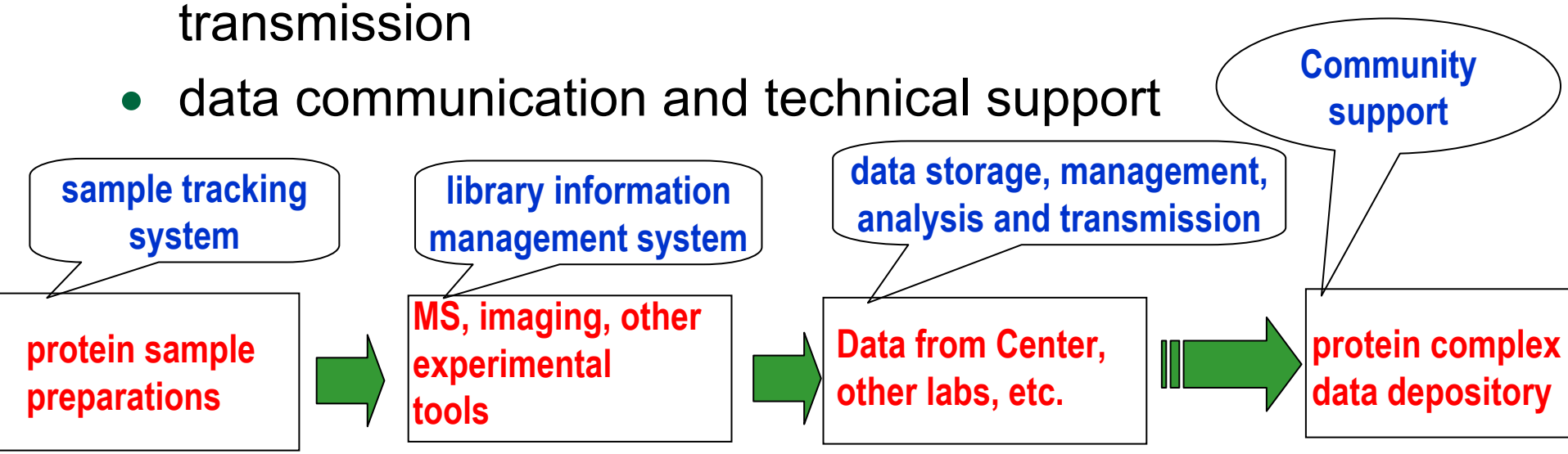
- Characterize protein complexes in isolation, within cells, and on cell surfaces/interfaces
- Employ multimodality approaches to molecular imaging—optical probes, molecular recognition force microscopy, afm/optical, $(optical)^n$
- Validate the composition of protein complexes
- Determine the location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

# Bioinformatics and Computing

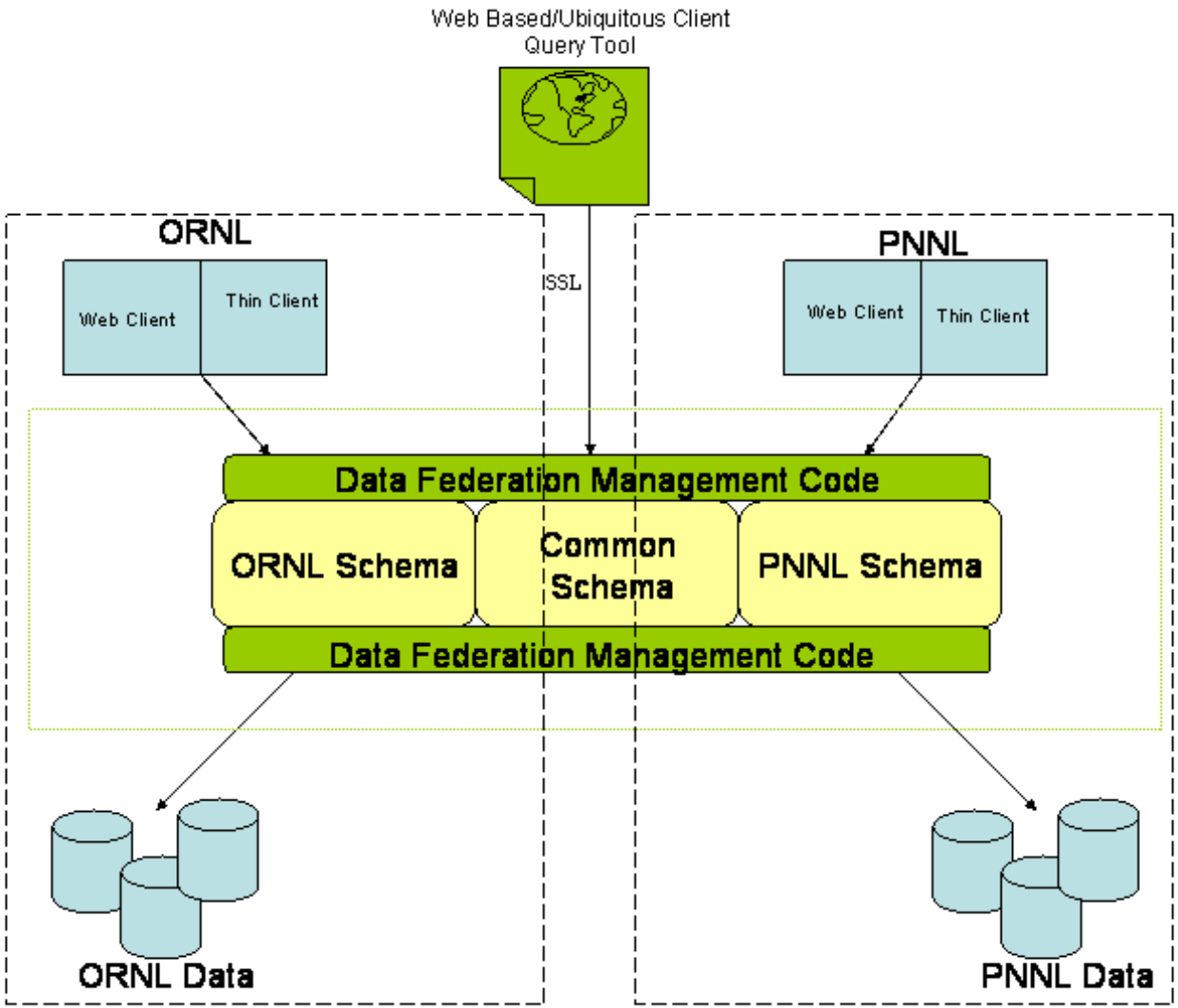
- **Short-term goals**
  - Create infrastructure for sample tracking, data collection and analysis
  - Improve tools for predicting and validating members of protein complexes
  - Build tools for interpreting MS data from cross-linked and modified proteins
- **Long-term goals**
  - Predict protein structures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



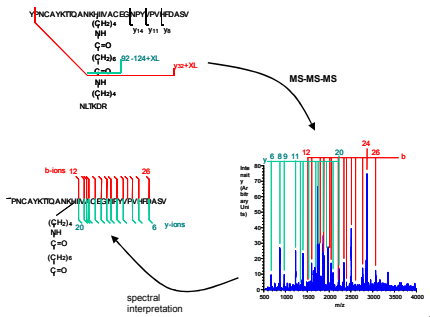
# GTL LIMS System Architecture



# Computational Characterization of Protein Complexes

Experimental conditions

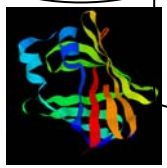
mass spec. data




genomic/proteomic databases

emerging technologies

PDB



structural genomics

data interpretation and modeling

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

structural characterization of protein complexes

protein interaction map construction

protein identification and protein complex identification

Center for Molecular & Cellular Systems

# Acknowledgements


- Research sponsored by Office of Biological and Environmental Research, U.S. Department of Energy.