2.0 GENOMES TO LIFE CENTER FOR MOLECULAR AND CELLULAR SYSTEMS

Abstract

Our long-term goal is to develop experimental strategies for the rapid identification of a substantial fraction of the multiprotein complexes in an organism. These strategies will be applicable to a wide range of organisms, allowing us to understand how complex formation is affected by growth conditions. Our *hypothesis* is that the identification of the molecular complexes within microorganisms, and dynamic changes within these complexes in response to growth conditions and environmental changes, will provide important insights into the metabolic regulatory strategies used by these organisms to adapt to environmental changes. We propose to implement a strategy to rapidly isolate and identify protein complexes from microbes with known DNA sequences, using existing state-of-the-art at both Oak Ridge National Laboratory (ORNL) and the Pacific Northwest National Laboratory (PNNL) as well as new technologies that will be developed in the four research subtasks included in the Project. Both ORNL and PNNL will contribute distinct, but complementary cell-free and *in vivo* approaches, which will allow comparison and identification of the most robust experimental approach for this task. The initial focus represented in this proposal will involve *Shewanella oneidensis* (*S. oneidensis*) *and Rhodopseudomonas palustris* (*R. palustris*), whose metabolisms are important in both understanding microbial energy production and environmental remediation. To accomplish these goals, we propose five specific aims: (1) Identify and Express Candidate Proteins Involved in Complex Formation, (2) Differentially Express Critical Protein Complexes, (3) Optimize Purification of Complexes Following Cellular Disruption and Fractionation, (4) Identify Proteins and Their Stoichiometries in Isolated Complexes, and (5) Integrate Data Management and Analysis with Characterization of Protein Complexes.

Principal Investigators:

ORNL:
Dr. Steve J. Kennel
Life Sciences Division
Oak Ridge National Laboratory
P. O. Box 2008, MS-6101
Oak Ridge, TN 37831-6101
Phone: 865/574-0825
Facsimilie: 865/576-7651
E-mail: kennelsj@ornl.gov

PNNL:
Dr. Thomas Squier
Division of Biological Sciences
Pacific Northwest National Lab
P.O. Box 999 / MS P7-53
Richland, WA 99352
Phone: 509/376-2218
Facsimilie: 509/376-1494
E-mail: Thomas.Squier@pnl.gov

2.2 Background and Significance

Monitoring alterations in the expression of specific cellular proteins in response to environmental stress or alterations in growth conditions has provided some insight with respect to the linkages between specific metabolic pathways. However, this approach is limited since individual proteins rarely carry out their functions in isolation. Rather, it is now understood that metabolic flux through specific pathways is greatly facilitated by the formation of protein complexes that minimize diffusional barriers that slow reaction rates. Furthermore, specific proteins frequently serve as modular components that function as subunits in more than one protein complex [Kumar and Snyder, 2002]. Thus, an understanding of cell function requires an understanding of the binding partners present in a protein complex that correlates with adaptive cellular responses to defined growth conditions. Therefore, the molecular basis for understanding energy balances and environmental challenges important to the goals of DOE requires the development of new strategies for identifying protein complexes in an organism under different growth conditions. Critical to this mission is the development of high-throughput methods. To accomplish these aims, we propose to initially identify the protein complexes in two organisms of fundamental interest to DOE, i.e., *S. oneidensis* and *R. palustris*. The genomes of these organisms have been sequenced and annotated, and thus the ORFs have been identified. Knowledge of all possible protein sequences from the genomic data permits the use of modern proteomic tools to define protein complexes, their functions, their regulation, and dynamic changes in protein-protein interactions as the bacteria undergo changes in growth conditions. Bioinformatics, cloning, tagging, protein expression, affinity purification, microscale separations, scFv production, molecular cross-linking, mass spectrometry, and imaging tools will be integrated for the common goal of understanding the biology of these complexes.

We propose to implement a strategy to isolate and identify protein complexes from microbes with known DNA sequences, using existing state-of-the-art technologies at both ORNL and PNNL and new technologies that will be developed as part of the research subprojects described elsewhere in this proposed project. Both laboratories will contribute distinct, but complementary cell-free and *in vivo* approaches, which will allow comparison and identification of the most robust approach for this task. The initial focus represented in this proposal will involve *S. oneidensis* and *R. palustris*, whose metabolisms are important in both understanding microbial energy production and environmental remediation. To accomplish these goals, we propose the following specific aims.

Specific Aim 1: Identify and Express Candidate Proteins Involved in Complex Formation. Genes of interest will be cloned into shuttle vectors for cell-free and *in vivo* expression following identification of candidate proteins that are expected to form complexes. Expression of cloned proteins that are "tagged" to facilitate purification will permit complex identification following cell disruption. Following immobilization on resins, these proteins will serve as "bait" in the purification of cellular binding partners. These expressed proteins will aid in the generation of single chain variable fragment (scFv) antibodies.

Specific Aim 2: Differentially Express Critical Protein Complexes. Growing microorganisms under variable conditions that emphasize different metabolic pathways will permit quantitative correlations to be made between protein complex formation and cell function. This will identify individual proteins that function as subunits in more than one complex, and elucidate the coordinate regulation of important protein complexes involving, for example, metabolite flux, protein synthesis and degradation, DNA replication and repair, transcription and translational activities, and antioxidant enzyme activities. An additional advantage of these measurements is that they will also test the completeness and specificity of complex identification, since different growth conditions will result in preferential expression of selected pathways of metabolism.

Specific Aim 3: Optimize Purification of Complexes Following Cellular Disruption and Fractionation. Affinity resins and high-throughput separation methods with minimal sample handling will be developed that permit the purification of protein complexes suitable for mass spectrometric identification. Two general approaches will be used: unique molecular tags engineered into the protein sequence and scFv antibodies reacting with native sequences. Complexes bound to resins will be eluted using mild conditions that maintain oligomeric interactions within complexes, and that will permit a reassessment of the specificity of the complex using a complementary method that may involve either cycling the complex back onto the stationary phase resin or the

use of an alternate entry point into the complex. Weakly associating complexes will be stabilized through cross-linking methods developed at ORNL.

Specific Aim 4: Identify Proteins and Their Stoichiometries in Isolated Complexes. Initial approaches for identifying isolated proteins within complexes will be based on both analysis of intact proteins and of peptides produced by proteolytic digestion using mass spectrometry (MS). Two approaches will be evaluated, both the peptide level accurate mass tag (AMT) methodology developed at PNNL and multi-dimensional LC-MS techniques at ORNL. Stoichiometries of proteins in complexes will be identified by whole protein analysis using SDS-PAGE and MS. Purified expression proteins will serve as standards to relate the observed signal to the relative abundance of the observed protein.

Specific Aim 5: Integrate Data Management and Analysis with Characterization of Protein Complexes. New data processing capabilities will be developed that permit the rapid on-line sharing of databases that, for example, provide transparent linkages between (a) sample data and their processing-status information, (b) mass spectrometry data, (c) interaction maps of identified complexes, and (d) detailed structural and functional information of identified protein complexes. The data will be stored locally at ORNL and PNNL and will be integrated following the development of web applications that will be accessible through the Center's web portal.

Strategic Plan: Taking advantage of the fully annotated genomes that permit identification of the open reading frames (ORFs) of all proteins in *Shewanella* and *Rhodopseudomonas*, two independent approaches will be investigated simultaneously to identify the optimal methods for complex identification. These approaches emphasize exogenous "tags" on proteins expressed using cell-free translation methods (PNNL) and endogenous "tags" on proteins expressed *in vivo* (ORNL). It is expected that the cell-free protein expression system will become fully functional within the first four months, permitting: (a) immobilization of tagged proteins on affinity columns, where they will function as "bait" in the capture of protein complexes and (b) use of isolated proteins in the high-throughput generation of scFv for affinity purification of protein complexes. The expression of tagged proteins *in vivo*, like antibody methods, will permit the capture of stable complexes that form during protein synthesis (following cell disruption) on affinity resins that might not bind to immobilized "bait" proteins. Furthermore, these complementary methods will provide increased confidence in the cellular relevance of isolated complexes, and will permit a comparison of the reliability of each method for subsequent development of high-throughput sample identification methodologies. Criteria that will be used in determining the most robust methods, which will then be standardized at both laboratories, include (a) demonstrated throughput (complexes per unit time), (b) data quality (reproducibility and consistency between different approaches), and (c) sensitivity (sample amount). During the second year optimal methods will be further developed and established at both laboratories, resulting during year three in the necessary rapid high-throughput methods for routine complex identification of more than 80% of all protein complexes in any microbe (whose genome has been sequenced and annotated) in time-periods considerably less than one year.

2.2.1 Rationale For Choice of Microorganisms

The large data base available from the genome sequence of *S. oneidensis* and *R. palustris* will facilitate the development of high-throughput proteomic methods, which will be capable of identifying the protein complexes in any microorganism. However, the choice of these microorganisms has the additional advantage in that these bacteria are important for understanding a wide range of bioremediation and energy cycles (see below). For both *S. oneidensis* and *R. palustris*, protein complex formation will be assessed under different culture conditions intended to stimulate or repress expression of the various enzyme complexes. This will test the qualitative and quantitative aspects of protein complex analysis.

S. oneidensis is a facultative anaerobic heterotrophic proteobacterium, which can colonize a wide range of environments because of the ability to use a wide variety of terminal electron acceptors, ranging from the common (oxygen, nitrate and nitrite) through the sulfur family (elemental sulfur, thiosulfate) to organics (fumarate, dimethylsulfoxide) and metals (iron, uranium). This ability to reduce metals is of particular importance, since the migration of heavy metal contaminants in the soil may be manipulated by controlled use of microbes such as *S. oneidensis*, which have the potential to function as biobarriers. The process of metal

reduction in *S. oneidensis* is unique in that the electrons are ferried from the cell's interior out through the periplasm to the outer membrane, where they can be transferred directly to the oxidized metal [Newman and Kolter, 2000]. However, little is currently known about the protein complexes required for the reduction of heavy metals or other electron acceptors by *Shewanella,* other than that electron transport through complex, multiheme cytochromes mediate this function. For this reason, the *Shewanella* Federation was formed within the DOE-OBER Microbial Cell Project (MCP) to attempt to understand metal reduction and other questions associated with remediation. Here we seek to support that effort by identifying critical protein complexes and their dynamic regulation in response to alterations in growth and environmental conditions. This task will be simplified by using *Shewanella*, since we already know the oligomeric structure of some critical protein complexes that are involved in energy metabolism. For example, succinate dehydrogenase (complex II in the respiratory chain) is a plasma membrane protein with four subunits (termed A, B, C, and D). Subunits A and B are hydrophilic, whereas subunits C and D are integral membrane proteins [Lancaster, 2002]. These and other known protein complexes (see **Table 2.2.1**) will serve to test and validate proposed proteomic methods of analysis.

Table 2.2.1. Listing of Known Protein Complexes in *S oneidensis* and *R. palustris*.

Complex Name	Organism	Cellular Location	Subunits	Metabolic Regulation	Reference
Thiosulfate oxidase	*R. palustris*	Soluble	> 3	sox	Friedrich et al., 2000
Glutamine Synthetase	*R. palustris*	Soluble	Not Determined	nif	Cohen and Yamasaki 2000
	S. oneidensis	Soluble	Not Determined	nif	Ikegami et al, 2000
Hydrogenase	*R. palustris*	Membrane	Many	hupT	Baron and Ferry 1989
	S. oneidensis	Membrane	3	hup	Dischert et al , 1999
Nitrogenase	*R. palustris*	Soluble	Many	nif	Elsen et al, 2000
Cytochrome c Oxidase	*R. palustris*	Membrane	Not Determined	Unknown	
	S. oneidensis	Membrane	Not Determined	Unknown	
Benzoate Degradation	*R. palustris*	Soluble	4	badH	Pelletier and Harwood, 2000
NADH Oxidoreductase	*R. palustris*	Membrane	Not Determined	Unknown	
	S. oneidensis	Membrane	14	Unknown	
GroELS	*R. palustris*	Soluble	2	Unknown	Hemmingsen et al., 1988
	S. oneidensis	Soluble	2	Unknown	
Fumarate Reductase / Succinate Dehydrogenase	*S. oneidensis*	Soluble	1	Unknown	Reid et al., 2000; Leys et al., 1999
	S. oneidensis	Membrane	4	Unknown	Tsapin et al., 2001
Fumarate Reductase	*R. palustris*	Membrane	4		Myers and Myers, 2000
Nitrate Reductase	*S. oneidensis*	Soluble	Not Determined	Unknown	
Iron Reductase	*S. oneidensis*	Membrane	Not Determined	Unknown	Beliaev et al., 2001; Field et al., 2000
TMAO/DMSO Reductase	*S. oneidensis / R. palustris*	Soluble	Not Determined	tor	Knablein et al., 1997

* sequence information on these complexes can be found at: www.genome.ad.jp/dbget-bin/w

R. palustris is a member of the non-sulfur purple bacteria that exhibits an extremely diverse metabolic functionality that permits utilization of almost every major class of energetic and assimilatory metabolism. Thus, like *S. oneidensis*, *R. palustris* can grow under a wide range of environmental conditions. It can function as a phototroph, heterotroph, autotroph, lithotroph, or facultative anaerobe while performing nitrogen fixation. Thus, the growth of *R. palustris* can be adapted to fix CO_2 or to use organic molecules as a carbon source. Similarly, N_2 can be fixed from the atmosphere, or NH_3 can be used as a nitrogen source for basic metabolism. Energy for these processes can be obtained from photosynthesis or from reduced substrates. In the course of these processes

the bacterium can degrade lignin and reduce various metals and oxyanions. The regulation of these processes and the protein complexes that perform the functions are of basic interest to the carbon, nitrogen, and energy cycles on earth. Significant work has been done on the induction and interactions of these enzyme systems [Joshi and Tabita, 1996; Qian and Tabita, 1997; Tichi et al., 2001], providing an appreciation for some critical complexes important to key metabolic pathways. When this bacterium is cultured under different conditions, the appropriate carbon, nitrogen and energy handling systems are induced. This single organism provides a variable pool of protein complexes that depend on the specific culture conditions. Thus, *R. palustris* is a good starting point on which to develop robust protocols for future high-throughput protein complex analysis for other organisms with less diverse metabolisms.

Our initial strategy will be to identify partially characterized protein complexes of known oligomeric proteins (e.g., succinate dehydrogenase) and critical metabolic enzymes as input points in complex purification, and to identify other members of these as well as other complexes. It is expected that binding partners should involve proteins in the same regulated pathway. For example, we will assess possible associations between soluble nitrogenase with a hydrogenase that will be dependent upon cellular growth conditions. Likewise, upon shifting growth conditions to benzoate, we might expect specific protein associations between elements of the benzoate degradation pathway.

2.2.2 Purification of Protein Complexes

A major challenge is to detect dynamic multiprotein complexes involving low-abundance proteins using a high-throughput methodology. An additional challenge is to detect membrane protein complexes following disruption of the membrane lattice. Standard MS methods have proven useful in identifying complexes of abundant proteins (present at 15 copies/cell), and suggest that the majority of expressed proteins undergo complex formation with partners [Gavin et al., 2002; Ho et al., 2002]. To date, these methods have involved molecular tags expressed on individual proteins, which permit purification of high-affinity complexes upon association with an antibody bound to an affinity matrix [Rigaut et al., 1999; Puig et al., 2001; Einhauer et al., 2001; Kumar and Snyder, 2002]. For example, Ho and colleagues used a FLAG epitope (DYKDDDDK) to purify protein complexes in yeast with a single step procedure, while Gavin and colleagues used a tandem affinity purification (TAP) method in which a large (20 kDa) tag was engineered into each protein [Ho et al, 2002; Gavin et al., 2002; Séraphin et al., 2002]. In the latter case, the affinity tag contained two sequences that permitted sequential purification involving protein A absorption to an IgG-matrix, proteolytic cleavage of the absorbed complex, and binding of the tagged protein to a calmodulin affinity column through a 26 amino acid calmodulin-binding peptide (CBP) (KRRWKKNFIAVSAANRFKKISSSGAL) [Rigaut et al., 1999]. The methodologies of tagging the proteins involved over-expression and isolation of proteins from cells grown at log-phase, and complexes were eluted following extensive washing of nonspecifically bound proteins from the affinity column. In control experiments, using the TAP method, protein recovery was estimated to be approximately 40% [Rigaut et al., 1999]. It is expected that the significant nonspecific protein association observed using this methodology is due to the highly charged epitope tags (see above) used for affinity purification, which results in false positives in the common screens used to date. Indeed, both methods discussed above "find a significant number of false-positive interactions, while failing to identify many known associations," suggesting that it "will be essential to integrate data from many different sources to obtain an accurate understanding of protein networks" [Kumar and Snyder, 2002]. Furthermore, there is minimal overlap between the sets of complexes identified. New and improved methods are needed to assure the reliability and completeness of complex identification. Initially, we propose to use proteins tagged with the GST peptide for affinity isolation, which may be supplemented with a second purification via a biotin tag introduced with crosslinking reagents (see **High Throughput Molecular Characterization subproject**). We will also isolate native proteins using scFv's as they become available (see below).

2.2.3 Identification of Isolated Proteins

Most protein characterization approaches now employ separation methodologies combined with some form of mass spectrometry (MS), typically applied after proteolytic digestion. A widely used approach involves tryptic digestion of the protein and subsequent MS analyis of individual peptides that are then fragmented and identified

using mass-to-charge ratio (*m/z*) measurements (i.e., MS/MS). Often MS/MS analysis of only one peptide is sufficient for protein identification. Individual proteins within a complex are identified following proteolytic cleavage (e.g., using trypsin) that produces tens to hundreds of peptides from each protein in the complex--perhaps 10^3 to 10^5 in total,depending upon complex size. This complex peptide mixture is then analyzed by capillary liquid chromatography (LC) combined directly with high performance mass spectrometry using commercially available quadrupole time-of-flight and quadrupole ion trap, and Fourier transform ion cyclotron resonance (FTICR) mass spectrometry.

The power of MS for protein identification derives from the specificity of mass measurements for either the intact peptides or their fragments after dissociation in MS/MS measurements. Thus, given sufficient mass measurement accuracy (MMA) it is possible in a bacterium to unequivocally identify a peptide sequence from the mass. Since each protein is likely to have multiple unique peptides, and additional constraints are obtained from prior fractionation steps or the use of LC elution times, it is possible to uniquely identify a protein from the peptide mass measurements in approximately 98% of all cases [Smith et al., 2002] when combined with knowledge provided by a single genome sequence and its predicted proteome. Thus, given sufficient MMA, a peptide mass measurement can usually be attributed to a single protein within the constraints provided by a single genome sequence and its predicted proteome. The limited MMA achievable with conventional MS technologies generally necessitates additional separation steps or the use of MS/MS methods to provide additional data to support protein identification. An alternative approach is to us AMTs (see below), which potentially obviate the routine need for MS/MS, and correspondingly reduces sample size requirements.

2.3 Preliminary Studies

We propose to build upon existing expertise with respect to both the microorganisms under investigation as well as state-of-the-art mass spectrometric methods to develop a high-throughput capability for the identification of protein complexes. Critical to an identification of protein complexes is the development and application of predictive algorithms and databases that permit the rational choice of proteins that are likely to form complexes. It is, therefore, possible to select proteins from the genomic data that are likely to form complexes involved in critical points of metabolism as starting points in defining the protein complexes in any particular bacterium, including *S. oneidensis* and *R. palustris*. Furthermore, since it is likely that the majority of expressed proteins form complexes, it will not be necessary to clone and express the entire genome to identify the full array of protein complexes. Following the choice of target proteins for cloning, it is necessary to design molecular tags that permit an expressed protein to be recognized by an affinity matrix for rapid purification. Our approaches will utilize mild conditions that retain binding interactions within a complex, but which minimize nonspecific protein binding that can interfere with the identification of unique members of a protein complex. We will use either an endogenous tag within the cloned protein, or alternatively exogenous binding reagents (e.g., scFvs) that recognize unique sequences within individual native proteins. Following isolation, we will identify the individual proteins within the complex and their relative stoichiometries. In many cases, protein identification involves the initial separation of individual protein components, using SDS-PAGE, and subsequent MS analysis. Alternatively, liquid chromatography (LC) can be used to separate individual proteins (or tryptic peptides), and with sufficient mass resolution it is possible to uniquely identify the protein components of a complex. The cellular relevance of an isolated complex will be verified with a complementary method. Examples include affinity isolation using an alternative binding partner in the complex as well as direct assessments of the strength and specificity of the binding interaction that can be measured following covalent attachment of fluorescent probes to individual proteins [Sun and Squier, 2000]. These methods will facilitate visualization of protein complexes within living cells (see **Molecular and Cellular Imaging subproject**).

2.3.1 Identification of Candidate Proteins for Cloning and Expression

2.3.1.1 Sequence and Annotation

The *S. oneidensis* genome is finished and the publication of the annotated sequence is imminent. In addition, ORNL has independently annotated this genome for the *Shewanella* Federation. ORNL is also responsible for the final annotation of the *R. palustris* genome; a manuscript describing the annotated sequence is in preparation.

Functional assignments of individual genes, phylogenetic relationships, and predicted operon structures are available to aid target selection.

2.3.1.2 Protein Tags

Affinity tags are routinely used to isolate protein complexes and have been used at ORNL to isolate the complex between cyclin E and other proteins involved in cell cycle control in mammalian cells. For example, cyclin E was tagged with GST and shown to form a complex with the CUL1 protein. Data in **Fig. 2.3.1.** demonstrates that the GST tag could be used to purify the cyclin E-CUL1 complex (lane 5). However, the typical construction protocol is inadequate for the high throughput processing of large gene sets, which precludes the use of traditional restriction enzyme based cloning protocols. As such, the *Shewanella* project is integrating PCR amplified ORFs from the *S. oneidensis* genome into a commercially available site-specific recombination vector. The ORF contained in this "donor" or entry plasmid can then be recombined, using the Cre/*loxP* site-specific recombination system, into a variety of "acceptor" or expression vectors placing the cloned gene under the control of different promoters fused in-frame to given affinity tags of choice. The ability to select positively for only the recombined expression vectors along with the site-specific nature of the transfer greatly reduces both the time and resources spent on clone verification as ≥ 95% of resulting transformants contain the correct gene sequence and reading frame. This recombinational cloning system is composed of a 38-kDa Cre recombinase that mediates recombination between two 34-bp *loxP* sites without the need for accessory factors and can occur in cell-free homogenates or *in vivo*. The *loxP* recombination sequence has not had, to date, a deleterious impact on the expressed protein and the simple addition of the *loxP* sequence allows any expression vector to be easily adapted for use with this system.

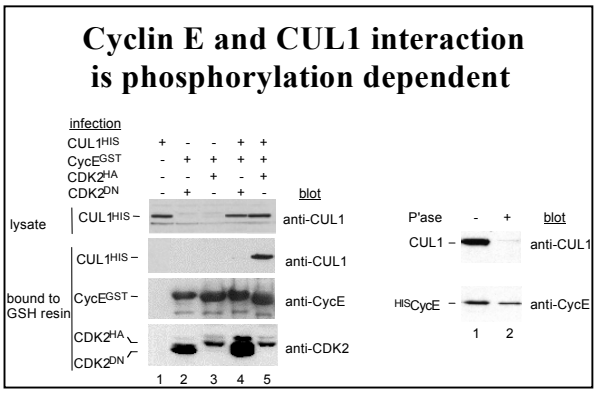

Fig. 2.3.1. Phosphorylation-dependent transient proteins in complex with cyclin E are resolved using SDS-PAGE following affinity purification.

2.3.1.3 Cell-free Translation

The generation of tagged proteins for complex identification requires the rapid cloning of the majority of *Shewanella* genes into vectors that can be used for cell-free and *in vivo* expression and complex identification. We propose to use a cell-free translation system as an effective means to obtain milligram quantities of proteins in a high-throughput manner (see **High Throughput Complex Processing subproject**). We have carried out cell-free protein translation previously [Kocarek et al., 2000], and are evaluating the Roche commercial cell-free transcription/translation system for high-throughput protein translation. We, furthermore, have extensive experience in the expression and purification of recombinant proteins, all with and without epitope tags, using traditional expression and purification methods (**Fig. 2.3.2**). Thus, sufficient expertise exists to accomplish these goals.

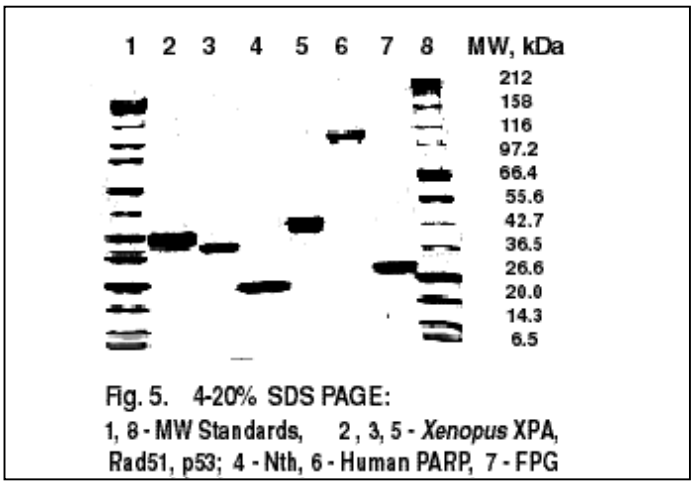

Fig. 2.3.2. Expression and purification of recombinant proteins without affinity tags.

2.3.2 Bacterial Cell Growth and Differential Expression of Protein Complexes

A wide variety of microbial cultures, including *Shewanella* strain MR-1, can be grown using the suite of standard bioreactors, continuous flow centrifuges, and analytical instruments needed for the controlled cultivation and

preparation of cells from which targeted protein complexes will be purified. Photobiosystems range from incubators to one liter bubble columns. A wide range of cultivation systems is available. These include continuous culture reactors (1, 5, 10, and 80 liter capacities), pilot scale fermenters (70 and 500 liter capacities), annular biofilm reactors with attached confocal microscopy, and extremophile reactors. Batch processing of grown cultures is routine, using large-scale harvesting capabilities that involve flow-through centrifugation to harvest one liter of cell culture per minute and tangential flow filtration units fitted with ultrafiltration membranes to purify outer membrane vesicles from Gram negative bacteria (e.g., *Shewanella*). This experience at bench-to-large-scale cell culture and fractionation methods provides a solid basis for the rapid growth and isolation of bacterial homogenates necessary for this project.

2.3.3 Complex Purification

Purification of complexes is necessary to reduce the number of peptides that must be identified and monitored. In all purification schemes, it is necessary to enlist stringent separation techniques to remove non-specifically bound proteins and yet use conditions mild enough to retain complex integrity. In this latter respect, the association between protein subunits within a complex can be stabilized through intermolecular cross-linking prior to purification. Critical to the identification of molecular complexes is the development of affinity purification techniques, which permit identification using MS [Yang et al., 1996].

2.3.3.1 Affinity Purification

Using scFv antibodies, we have previously developed methodologies that permit the rapid isolation of protein complexes from whole cellular homogenates involving Ku86 and Ku70 [Siegel et al., 2000; see **Fig. 2.3.3.**). Thus,

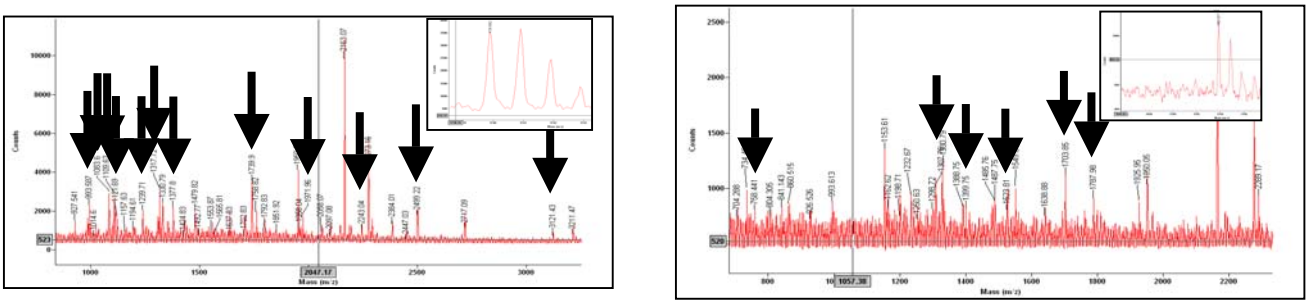

Fig. 2.3.3. MALDI-TOF mass spectra from anti-Ku86 scFv immunoprecipitation. (A) Ku86 spectrum (B) Ku70 spectrum. The monoisotopic masses (insets) resulting from the enzymatic digests were matched against the SwissProt database. Arrows denote peptides that were matched to either Ku86 or Ku70 with a 40 PPM mass tolerance.

the development of high-throughput methods to generate scFv antibodies to the majority of bacterial ORFs will permit the implementation of this methodology to isolate protein complexes rapidly from cellular homogenates using a variety of entry points to ensure that isolated complexes are physiologically important. This technology is readily transferred to include other affinity strategies, and is scalable to permit a high-throughput isolation and on-line identification of protein complexes, as fully described in the **High Throughput Complex Processing subproject**. Briefly, the affinity matrix will be loaded into minicolumns and exposed to concentrated cellular fractions to permit saturation of the binding resin using optical detection to monitor binding. Alternatively, batch methods involving incubation of the affinity matrix with cellular homogenates in microtiter plates can be automated, and offers the potential to enhance the diffusion-limited binding of protein complexes to the matrix. Following purification, the complex can be eluted using standard methods involving either formic acid (pH 2) or acetonitrile, which are compatible with subsequent mass spectrometric protein identification. Under these harsh conditions the purified complex will be disrupted, and cannot be re-purified by the tandem affinity procedures often used to reduce nonspecific binding to obtain accurate identification of proteins in a complex. Therefore, it is expected that methods will be developed that permit release of the intact complex using mild conditions involving the reduction of disulfide bonds to release antibodies and isolated complexes from the column. For example, scFv antibodies can be linked to biotin by disulfide bonds and, following complex isolation, can be eluted from the stationary phase following addition of a reducing reagent (e.g., 50 mM DTT). The complex can subsequently be

re-purified to reduce the amount of nonspecific protein associations through steps involving the recycling onto the same column, or could use, for example, a commercially available (Pierce) monomeric avidin affinity column that binds the biotin in a cross-linker (see below).

2.3.3.2 Cross-linking

Cross-linking of proteins in a complex can be done before or after their purification. If cross-linking is done before complex isolation, it is possible that loosely attached proteins might be identified that otherwise could be lost during purification. The proteins can then be proteolytically digested and cross-linked peptides can be purified and identified using MS [Rappsilber et al, 2000]. This approach, when coupled with the use of different length cross-linkers, can also help define the spatial relationships of the proteins in the complex. Preliminary data or MS analysis of cross-linked peptides are presented in the **Bioinformatics and Computing subproject**. Preliminary data for synthesis of biotinylated crosslinking reagents are presented in the **High Throughput Molecular Characterization subproject.** In general, two N-hydroxysuccinimide linker groups are separated by spacer arms of various lengths. A biotin molecule is incorporated near the middle of the spacer arms. These reagents will be used to cross-link isolated protein complexes following their isolation or within a cellular homogenate. The biotin will subsequently be used to "fish out" the linked proteins and the linked peptide pairs after proteolytic digestion, aiding in sample preparation efficiency.

2.3.4 Protein Separation and Identification

Routine separation methods involving either SDS-PAGE or liquid chromatography are used to separate individual proteins or proteolytic fragments. However, in either case protein identification requires the use of MS to identify the masses of unique proteolytic fragments. Additional information regarding the specificity of an identified binding interaction is obtained from a consideration of the binding stoichiometries of individual protein components, since nonspecifically bound proteins are not likely to be present in a similar stoichiometry to the tagged protein used for complex isolation.

2.3.4.1 SDS-PAGE Separation

SDS-PAGE is commonly used to separate individual proteins following complex isolation, permitting in-gel proteolytic digestion and mass spectrometric identification of individual proteins in the complex. Following protein staining, a densitometric analysis of the gel bands provides additional information regarding the stoichiometries of individual proteins in the complex. This latter assessment is critical, since distinctions must be made between complex members that are present in stoichiometric ratios from those present in lesser amounts. Bands staining at or near the same intensity (correcting for mass differences) are likely to be present as integral members of the complex, whereas components that stain lighter may be contaminants or accessory molecules. Additional resolution of proteins in large complexes as well as changes in the expression levels of proteins within the cell can be resolved using 2-D SDS-PAGE [O'Farrell, 1975; Anderson et al., 1985]. This approach was, in fact, critical to the identification by ORNL of increases in the expression levels of a key metabolic enzyme (i.e., flavocytochrome *c* fumarate reductase) in *S. oneidensis* when grown under anaerobic conditions [Beliaev et al., 2002].

2.3.4.2 LC-MS Protein Identification

Following complex isolation, MS will be used to identify the proteins and their relative stoichiometries in protein complexes. Two different approaches will be assessed, one that employs AMTs and the other, multi-dimensional LC-MS. Conventionally, the generation of AMTs involves a two-stage process (**Fig. 2.3.4.**). In the first step, the proteome is proteolytically digested (typically using trypsin) and analyzed by capillary LC-MS/MS. During the separation, peaks from each spectrum are sequentially selected for analysis based upon pre-defined criteria designed to minimize repeated analysis of the same species. Ion trap-MS/MS measurements yield "potential mass tags" (PMTs) that are subsequently validated as AMTs if the predicted peptide's accurate mass is observed using FTICR at an equivalent elution time.

In the initial proteomic efforts with *D. radiodurans* at PNNL, a large fraction of the PMTs were generated using a "shotgun" approach with conventional ion trap mass spectrometers to conduct MS/MS measurements, and peptides were tentatively identified (i.e. designated as PMTs) using the search program SEQUEST. Conventionally, many MS/MS spectra and search results are manually examined to evaluate both spectrum quality and the ranking of peptide scores, generally resulting in discarding a substantial fraction of tentatively identified peptides. Our approach using highly accurate mass measurements provides an additional and high-quality "test" for a tentative peptide, and significantly increases the confidence of protein identifications [Smith et al., 2002].

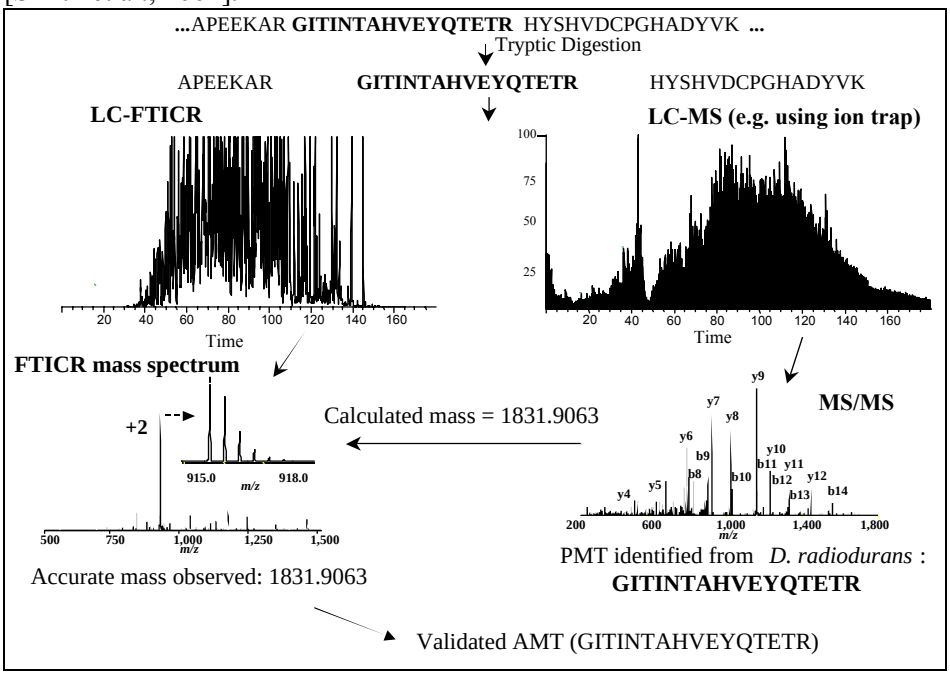

Fig. 2.3.4. Experimental steps involved in establishing an AMT In the first stage, peptides from a tryptic digestion are analyzed by LC-MS/MS and tentatively identified as potential mass tags (PMTs). The analysis is also conducted using a high-field FTICR mass spectrometer. An AMT is established when a peptide is detected eluting at the same time and corresponding to the calculated mass of the PMT is measured. This peptide then functions as a biomarker to identify the protein in subsequent studies of protein complexes, and results in much greater throughput.

Initial demonstration of the use of AMTs for proteomics: The use of AMTs has been initially demonstrated for the small radiation resistant, prokaryotic organism *Deinococcus radiodurans*, whose ~3.1 Mbase genome is *predicted* to code for 3187 proteins. Our calculations show an ideal tryptic digest of all proteins would yield 60,068 peptides between 500 to 4000 Da, of which ~51% would be unique at 1 ppm MMA, potentially affording coverage of >99% of all proteins. Analyses of *D. radiodurans* cultured under a number of different conditions, typically resulted in the detection of 20,000 to 60,000 peptides. Using capillary LC-MS/MS, we generated >9000 peptide PMTs, of which 6,997 were validated as AMTs. These AMTs allowed identification of 1,910 predicted proteins, confirming ~61% of the predicted proteins and spanning every category of predicted protein function from the annotated genome and corresponding to the largest coverage yet obtained for an organism of this size. The *D. radiodurans* proteins observed included a majority of the predicted proteins in most of the functional categories defined by TIGR, including 88% of the proteins associated with protein synthesis, 78% with transcription, and 80% of the predicted proteins involved in glycolysis and the pentose phosphate shunt.

2.3.4.3 Two-dimensional chromatography/MS

We have demonstrated multi-dimension chromatography interfaced with electrospray mass spectrometry in several different configurations at ORNL, as described in the attached manuscript on characterization of the proteome of *S. oneidensis* [VerBerkmoes et. al., 2002; see Appendix of the **High Throughput Molecular Characterization subproject**]. The first configuration involves off-line anion exchange FPLC fractionation of cell lysates. FPLC fractions were divided before two subsequent analytical paths: (1) The "bottom-up" path, in which all proteins in the fraction were digested by trypsin, and the resulting peptide mixture subjected to reverse-phase HPLC interfaced with electrospray/quadrupole ion trap tandem mass spectrometry (LC/MS/MS), and (2) the "top-down" path [Kelleher et al., 1999], in which the fraction was dialyzed to remove low-molecular-mass contaminants, and then intact protein molecular masses were measured to high accuracy using electrospray Fourier transform mass spectrometry. Computer-based and manual data interpretation techniques then allowed

identification of proteins from the two complementary sets of mass spectrometry data. The second multidimensional chromatography configuration combines two separations dimensions in a single system, as described for peptide mixtures by the Yates group [Washburn et. al., 2001; Wolters et al., 2001]. Serially connecting commercially-available cation-exchange and reverse-phase columns provides the required orthogonal separation without specialized equipment for fabricating columns. This LC/LC-MS/MS configuration allowed identification of over 400 proteins in a single multiple-hour run. Using these different multidimensional chromatographic configurations, a total of over 850 proteins, including 43 hypotheticals, from virtually every functional class were identified from *S. oneidensis*. In addition, the intact protein molecular mass measurements provided important information on a subset of these proteins, in particular identifying a previously unknown 28-amino acid signal peptide on a putative periplasmic protein from an ABC transport system.

2.4 Research Design and Methods

The approach we propose for the isolation and identification of protein complexes is outlined below in five specific aims. We have designed these experiments to allow comparison of results from various techniques, both to evaluate the efficacy of the techniques and to validate the identities of the complexes. In addition, we will validate characterization of protein complexes by comparison with current "benchmark" techniques. For example, we will use denaturing and non-denaturing gels to confirm the identities of the complexes and compare stoichiometries obtained by MS techniques with those obtained by optical density measurements on gel-separated complexes. Although not specifically described here, as imaging technologies are developed in the **Molecular and Cellular Imaging subproject,** these tools will be incorporated into the Center to aid in characterizing protein complexes and assist in eludicidating the biological function of the complexes.

2.4.1 *Specific Aim 1*. Identify and Express Candidate Proteins Involved in Complex Formation.

This aim seeks to use the ORFs identified in the recently annotated genomes of *S. oneidensis* and *R. palustris* for *in vivo* and cell-free expression of individual tagged proteins. The emphasis of this specific aim on *S. oneidensis* and *R. palustris* reflects their significance to DOE in understanding energy metaboism and bioremediation. These experiments require that individual ORFs can be rapidly cloned into expression vectors such as the Cre/loxP site-specific recombination vector (commercially available from Invitrogen). The expression vectors will then be used for the cell-free translation of up to 200 cDNAs per day. Soluble translated proteins will be immobilized individually on a stationary phase and used as "bait" to isolate complexes. Alternatively, *in vivo* expression of tagged proteins will permit complexes to form during protein folding, and will permit their subsequent purification. Thus, in conjunction with the **High Throughput Complex Processing subproject**, we will develop affinity matrices that permit the high-throughput identification of protein complexes using automated renewable microcolumns.

2.4.1.1 Identification of Targeted Proteins for Complex Identification

We will develop a computational capability for predicting protein-protein interaction maps in our target genomes (initially, *S. oneidensis and R. palustris*, and later other DOE-relevant genomes), based on publicly available information. This information may include the following:

Protein-protein Interaction Information: The classical "two-hybrid system" allows genome-wide characterization of protein networks by identifying all detectable binary interactions between proteins encoded by the genome subject to the caveats discussed in Background. For the bacterial genomes that we are going to study, two-hybrid data may be limited. Nevertheless, the yeast interaction map may provide a useful reference template when constructing maps for our target genomes through homology searches. In addition, databases like DIP [Xenarios et al., 2002] and BIND [Bader et al., 2001] also contain protein interaction data derived from additional experimental approaches, such as mass spectrometry [Ho et al., 2002].

Information of Conserved Genetic Neighborhood: Proteins within a conserved genetic neighborhood probably interact with each other to form complexes. Identification of operons or "conserved gene contexts" can provide clues about which set of proteins may form a complex. Identification of operons is often done through a combination of discovering conserved

gene context across multiple related genomes and the sequential arrangement of these genes in the genomic sequence. There are a number of existing algorithms for identification of operons [Lathe et al., 2000].

Information from Gene Expression Data: In prokaryotic bacterial genomes, co-expressed genes may represent co-regulated genes. Hence identification of co-expressed genes could provide an indication of interacting proteins and protein complexes. Co-expressed genes can probably be identified through gene clustering based on their expression profiles. We have developed a gene expression data analysis and clustering software for this purpose, called EXCAVATOR [Xu et al., 2001, Xu et al., in press].

Information of Co-occurrence of Genes in Genomes: The phylogenetic profiling approach [Gaasterland and Ragan, 1998; Pellegrini et al., 1999] is based on the assumption that proteins that function together in a pathway or structural complex are likely to evolve in a correlated fashion. During evolution, all such functionally linked proteins tend to be either preserved or eliminated in a new species. A phylogenetic profile can be used to represent the presence or absence of a protein in every known genome. It has been shown that proteins having similar profiles tend to be functionally linked. The method of phylogenetic profiling will be used to establish the probability of two proteins interacting with each other and to predict the function of uncharacterized proteins.

Information from Gene Fusion: It is observed that some interacting proteins have homologues in another organism fused into a single protein chain. Identifying proteins that exist as one single chain in one genome and exist as separate entities in other genomes can help identify possible interacting proteins. The "Rosetta stone" method [Marcotte et. al., 1999] predicts interacting proteins, based on such observations. This method, although limited by the relative infrequency of fusion events, is highly sensitive. A single fusion event is often sufficient to define all derived single domains as potentially interacting and hence functionally linked. We will use the protein-pair information together with known pathway information to help annotate protein functions.

Intrinsic Disorder in Proteins. Some proteins are intrinsically unstructured and remain so until binding occurs with an appropriate partner [Dyson and Wright, 2002]. Furthermore, unstructured regions in proteins are often associated with binding interfaces [Garner et al., 1999; Dunker et al., 2001a; Dunker et al., 2001b]. For example, proteins likely to be involved as 'nodes' in signaling cascades are likely to contain substantial intrinsic disorder; *e.g.,* the anti-oncogene p53 is nearly 2/3 disordered. Dunker and coworkers have developed a program to predict disorder, and its utility has been experimentally verified [Iakoucheva et al., 2001]. An initial analysis of *Shewanella*'s genome predicts that ~20% of the proteins contain disordered regions consisting of 40 (or more) consecutive amino acids. These proteins will serve as initial targets for the investigation of protein complexes, and include transcription factors, putative resolvases, numerous proteins of unknown function, and several proteins assigned to the formate dehydrogenase cluster.

2.4.1.2 Choice of Cloning and Expression Vector

The need to perform high throughput processing of large gene sets precludes the use of traditional restriction enzyme based cloning protocols [Davison, 2002]. As such, we propose to integrate PCR amplified ORFs into a commercially available site-specific recombination vector. A version of the pUni vector is available that is adapted to topoisomerase-mediated cloning. The ORF contained in this "donor" or entry plasmid can then be recombined, using the Cre/*loxP* site-specific recombination system [Liu et al., 1998], into a variety of "acceptor" or expression vectors placing the cloned gene under the control of different promoters and including sequences coding for affinity tags. The ability to select for only the recombined expression vectors and the site-specific nature of the transfer greatly reduces both the time and money spent on clone verification as ≥ 95% of resulting transformants contain the correct gene sequence and reading frame. This recombinational cloning system is composed of a 38-kDa Cre recombinase that mediates recombination between two 34-bp *loxP* sites without the need for accessory factors and can occur *in vitro* or *in vivo*. The *loxP* recombination sequence has not had, to date, a deleterious impact on the expressed protein and the simple addition of the *loxP* sequence allows any expression vector to be easily adapted for use with this system. The recombinant expression vector can then be used as a shuttle vector, using the R6Kγ broad host range origin to sustain the plasmid in *pir*-adapted *S. oneidensis* or *R. palustris*. Common *E. coli* promoters, such as p*lac* or p*BAD,* are expected to be effective for *in vivo* expression in either *S. oneidensis* or *R. palustris,* and native promoters can be utilized as well. Suicide vectors are also available for the generation of chromosomal integration/displacement constructs. Construction of a comprehensive library of full-length genes from *S. oniedensis* is being undertaken at ORNL for the *Shewanella* Federation, using the pUni vector system. This resource can be leveraged to produce necessary "acceptor" expression vectors for cell-free and *in vivo* expression (below). Constitutive expression from recombinant

promoters as well as native promoters will be used in addition to inducible system such as p*lac* or p*BAD*. Targeted *R. palustris* genes will be introduced into the pUni vector from PCR amplified ORFs. High throughput processing using standard pipetting robots will be used for cloning and vector conversion. Constructs will be propagated in *E. coli* hosts prior to transfer to native hosts (by bacterial conjugation or electroporation) or for use in cell-free translation systems. The sequence of constructs will be verified by automated fluorescent DNA sequencing as necessary.

Some initial experiments will also over-express plasmid cDNA in transduced bacteria. The cDNA of interest will be cloned in to the pMG103 vector [Inui et al., 2000], which has been modified to contain the GST tag at the carboxyl-terminal end of the polyprotein coding sequence, driven by the plac promoter. GST fusion vectors are part of an integrated system for the expression, purification, and detection of fusion proteins. The pGEX plasmids (for example pGEX4T1 and pGEX3X available from Amersham BioSciences Inc.) provide a multiple cloning site for in-frame cloning a gene of interest to the carboxyl-terminal of the GST from the parasite *Shistosoma japonicum* [Smith and Johnson, 1988]. The plasmids are designed for inducible, high-level intracellular expression of genes or gene fragments as fusions with GST. This 26 kDa enzyme binds reversibly and with high-affinity to glutathione-containing affinity matrices. Since isoforms of GST are not normally found in bacterial cells, fusion proteins are easily purified by affinity chromatography.

2.4.1.3 *In vivo* Protein Expression

Expression constructs will be introduced into *S. oneidensis* or *R. palustris* hosts by electroporation, with selection for kanamycin resistance [Inui et al., 2000]. Transformants will be screened for production of the desired fusion protein. The target proteins will be expressed from relevant univector plasmids, using, for example, His-tags and/or GST-tags. Cultures of stably expressing clones will be selected and grown to log phase prior to disruption using a bead mill or by detergent/chaotrope extraction (see below). Production of the desired fusion protein will be verified by SDS-PAGE or by MS, as appropriate. As noted above, several control options may need to be investigated for plasmid-based expression in particular. Constitutive vs induced levels of expression may influence rates or states of assembly, so a careful comparison with native baselines will be carried out. In addition, N-terminal vs C-terminal fusions can have very different fates, as can different fusion moieties.

2.4.1.4 Cell-free Protein Expression

High-throughput methods are proposed to achieve the cell-free protein synthesis of 200 different bacterial ORFs per day (fully described in the **High Throughput Complex Processing subproject**), using methodology recently learned by Dr. E. J. Ackerman at PNNL and developed at RIKEN Institute in Yokohama, Japan. One of the essential components of cell-free translation extracts is the preparation of the *E. coli* S30 extracts [Pratt, 1984]. The *E. coli* strain is commercially available BL21 from Stratagene, and growth media is standard 2XTY. It is essential to harvest the cells in early log-phase, and to chill the culture flasks immediately, and then complete all subsequent processing steps at 4 °C under conditions designed to minimize introduction of RNases. Cell disruption is a crucial step, and optimal cell extracts are prepared by breaking recently-grown cells in a suspension of glass beads in a commercially-available Hitachi instrument. After breakage, the cells are immediately centrifuged and then dialyzed under conditions that minimize dilution of the S30 extract. The extracts can be stored in liquid nitrogen for several months with no appreciable loss-of-function.

Routine preparation of DNA for cell-free protein translation involves standard affinity columns from Qiagen. It is essential to produce highly concentrated T7 RNA polymerase. Since no suitable commercial source is available, enzyme will be purified with a clone provided by the Studier laboratory at Brookhaven National Laboratory. T7 polymerase, creatine-kinase, S30 extract, amino acids, DNA template, and buffer are combined and the coupled transcription/translation reaction incubates for approximately 8 hours. A tag incorporated onto the expressed protein is used for protein purification, in analogy to the proposed methods for protein complex isolation (see Specific Aim 3). High-throughput expression/purification is readily achievable using standardized Beckman robots with Qiagen reagent. Without optimizing purification conditions, at least 100 soluble proteins/day are produced in quantities of 2 mg each at RIKEN. This represents a 50% success rate in terms of attempted gene translations. Failures are overwhelmingly attributed to precipitated protein, rather than failed transcription or

failed translation reactions, indicating that altered translation conditions will produce, at most, 100 proteins per day.

2.4.2 *Specific Aim 2*. Bacterial Cell Growth and Differential Expression of Protein Complexes.

The goal of this aim is to test the hypothesis that protein complexes vary in response to changing metabolic needs of an organism, in a way that dictates the cellular response to a wide variety of growth conditions. To accomplish this goal, *S. oneidensis* and *R. palustris* will be grown under defined culture conditions that favor specific metabolic pathways. It is anticipated that families of different protein complexes form in response to an organism's energy needs, and that identification of these complexes will provide an understanding and predictive capability to manipulate bacterial metabolism for DOE goals involving energy production and bioremediation. Initial measurements will emphasize the detection of known protein complexes, which will provide a test of the cellular relevance and completeness of complex identification from cellular homogenates. Subsequent experiments will then be able to develop a fundamental understanding of metabolic regulatory mechanisms through the identification of protein complexes under defined growth conditions.

2.4.2.1 *S. oneidensis* strain MR-1

Using a high-throughput approach we propose to identify not only the structural linkages between important respiratory proteins, but also complex formation involving coordinated cellular activities involving, for example, DNA replication, transcription and translational activities associated with protein expression, and RNA and protein degradation. The latter activities are of particular interest, since the down-regulation of unwanted protein complexes is critical to the shift in metabolism that allows these organisms to quickly adapt to and thrive in diverse environmental situations. Therefore, additional protein complexes that we will target will involve the degradosome (involved in RNA degradation) and proteasome (involved in protein degradation). These latter complexes form multiple cellular associations that modify their activities. In the case of the degradosome, it is expected from earlier measurements in *E. coli* that the functional regulation of this protein complex will involve reversible associations with scaffolding proteins that modify rates of RNA degradation [Liou et al., 2001]. It is, therefore, apparent that high-throughput methods that identify time-dependent changes in complex formation under a range of growth conditions will provide an important level of insight into the mechanisms underlying alternate energy pathways and the potential to predict and modulate bacterial growth.

The initial growth conditions selected for specific complex expression will be to culture *S. oneidensis* strain MR-1 in a chemically-defined medium with lactate as the limiting nutrient and fumarate as the anaerobic electron acceptor. We will maintain low biomass (50 mg/L) intentionally to avoid or minimize cell/cell interactions. Aerobic growth on lactate will be used as a fumarate reductase repressed comparison. We will stimulate hydrogenase expression levels by growth using hydrogen as the sole energy source and either fumarate, nitrate, or ferric iron as electron acceptors. Under these conditions the fumarate reductase protein complex should be isolated from cellular homogenates. Basal medium and cultivation parameters involving minimal medium will be expanded to generate protein complex samples needed to fulfill the scientific objectives of the current proposal. Additional growth conditions, including the list of electron donors, electron acceptors, and compounds that will serve as the carbon source and limiting nutrients in continuous cultures appear in **Table 2.4.1**. These experiments will provide quantitative correlations between the formation of families of protein complexes that are coordinately regulated under different growth conditions.

2.4.2.2 *R. palustris*

The robust metabolic capabilities of this organism suggest an exquisite regulation of protein complex formation that will be under the control of specific growth conditions. Indeed, fourteen different growth states are possible. Initial experiments will emphasize the baseline condition (identified as RP-A), which will result in a robust anaerobic growth condition on light with malate (30-50 mM), carbon dioxide and ammonium sulfate (30mM). Photoheteroautotrophic growth under similar conditions provided 15 g biomass/d/L over a five day diurnal period with production of hydrogen and biopolymer [Carlozzi and Sacchi, 2001]. We will utilize several 1-2 L columnar glass bioreactors and operate under continuous light with a slow sparge of 95% N_2 / 5% CO_2. The initial

experiments will be batch growth. Later we will utilize daily sequential harvest fed-batch and ultimately develop continuous growth conditions (similar to the progress already achieved for *Shewanella*). Continuous cultivation is ultimately desired to create a dynamic steady state for the microbial population so that total protein concentrations in the proteome should be stable. This baseline batch cultivation (RP-A) will give sufficient biomass for the testing of the protein purification protocols. This should give strong expression of fumarate reductase and hydrogenase and repress nitrogenase. Other complexes (such as chromatophores) will be expressed under light but are not the initial targets. We will measure biomass using optical absorbance of the culture and dry weight.

Table 2.4.1. Electron donor, carbon source, and electron acceptor matrix to stimulate production of target protein complexes in *S. oneidensis* strain MR-1.

Target Complex	Energy Source	Carbon source	Terminal Electron Acceptor
glutamine synthase	Lactate	Lactate	Oxygen
nitrate reductase	Lactate	Lactate	Nitrate
fumarate reductase / succinate dehydrogenase	Lactate	Lactate	Fumarate
iron reductase/cytochromes	Lactate	Lactate	Ferric citrate
TMAO reductase	Lactate	Lactate	TMAO
Hydrogenase/ nitrate reductase	H_2	Malate	Nitrate
fumarate reductase / succinate dehydrogenase	H_2	Malate	Fumarate
iron reductase/cytochromes	H_2	Malate	Ferric citrate
TMAO reductase	H_2	Malate	TMAO

Two additional growth conditions will be also used once the initial protocols are developed. One will feature some of the most interesting metabolism of *R. palustris*. For anaerobic photoautotrophic growth with nitrogen fixation, the cultivation will omit the organic acid malate and the ammonium sulfate and change the atmosphere to 5% CO_2 / 95% H_2. We will utilize the same columnar photobioreactors under constant light. This condition should stimulate nitrogenase and hydrogenase, while decreasing but not eliminating fumarate reductase. The third condition is intended to feature more common aerobic chemohetrotrophic respiration and in some ways to function as a negative control. *R. palustris* will be grown under dark, aerated conditions on malate and ammonium sulfate. This should repress nitrogenase, hydrogenase and express fumarate reductase. Paired cultivations of the three conditions will be used to study of induction of the complexes. The photobioreactor cultivations will be initiated at ORNL with techniques shared with PNNL.

2.4.3 *Specific Aim 3*. Purify Protein Complexes Following Cellular Disruption and Fractionation.

This aim proposes to isolate protein complexes following cellular disruption. It is, therefore, critical that cell extracts and subcellular fractions are prepared in a manner that preserves protein complexes so that they can be characterized by MS. During cell disruption and handling numerous opportunities are available that promote degradation and disruption of physiologically relevant complexes as well as the formation of non-natural associations [Séraphin et al., 2002]. For example, addition of salts or detergents to disrupt cells can dissociate weak protein associations. Likewise, long procedures can result in protein denaturation, proteolysis, or aggregation – which can promote nonspecific protein association. For these reasons, methods for cell disruption and complex purification will emphasize mild conditions that rapidly disrupt cells and sub-fractionate different cellular components under conditions that approximate cellular conditions. Initial measurements will, therefore, involve disruption of the cell wall in *Shewanella* by enzymatic or chemical treatment, and separation of cell fractions by centrifugation [Krause and Nealson, 1997; Johansen et al., 1997). Alternatively, cell disruption may involve procedures involving direct physical disruption using, for example, a bead mill as indicated in Specific Aim 1 and further described in the **High Throughput Complex Processing subproject**. These methods are in contrast to routine methods that have been employed to assess the proteome in *Shewanella*, which typically involve cell membrane disruption using 9 M urea, 1% Triton X-100, or the addition of other detergents such as Brij [Myers and Myers, 1992; Krause and Nealson, 1997; Thompson et al., 2002; Beliaev et al., 2002]. In the latter experiments it is expected that some protein complexes will be disrupted.

Our strategy for complex purification emphasizes small and uncharged affinity tags and engineered scFv antibodies. Additionally, sample identification will require that the eluting buffer must be compatible with trypsin digestion and MS. Our methods will emphasize capture strategies that rely on the release of bound complexes under mild reducing conditions that will preserve protein complexes and that permit tandem affinity purification (TAP) methods. In this manner, false positives that result from nonspecific protein absorption and cross-reactivity with the matrix and/or isolated complexes will be minimized. In some cases, cross-linking strategies will be used to stabilize bound complexes, emphasizing bifunctional chemicals that include a molecular tag (e.g., biotin) on the intervening spacer arm. Further purification of these complexes can then take advantage of the molecular tag.

2.4.3.1 Cell Breakage and Fractionation

Maintenance of protein complexes often requires that cell disruption maintain the high protein concentrations typically observed inside cells, which is estimated to be approximately 200 mg/mL [Ellis, 2001]. Variation of lysis conditions (e.g., salt concentrations) may alter the composition of protein complexes and can provide insight with respect to the cellular location (e.g., peripheral membrane protein) or physical interaction between individual proteins. For these studies, the outer cell wall/membrane will be disrupted following incubation with lysozyme/EDTA in the presence of protease inhibitors (e.g., 100 μg/mL PMSF and 10 μg/mL of leupeptin). Rinsing this preparation with isotonic sucrose solubilizes periplasmic proteins (10-15% of the protein) and centrifugation allows separation of spheroplasts. The spheroplasts are then ruptured using a French press, and ultracentrifugation (200,000 x g) will be used to separate cytoplasmic proteins (supernatant) from inner and outer membranes. Since the cytoplasm is not significantly diluted, native protein complexes are retained. The inner and outer membrane fractions can be further separated by density gradient centrifugation.

The methods described above have the advantage of using subcellular isolation to prevent cross contamination between cellular components, but have the disadvantage of using several steps to process the sample; thus, it will be more difficult to automate and adapt to high throughput analysis. One way around this difficulty is to rupture the intact cells and analyze the cytosol/periplasmic protein complexes. One promising approach may involve osmotic shock or the addition of the cationic peptide protamine, which has been shown to induce the quantitative efflux of cytosolic proteins into the medium using a range of different Gram negative bacteria, including *S. oneidensis* [Johansen et al., 1997].

Membrane protein complexes can be isolated by subcellular fractionation following cell disruption and protein cross-linking. Membrane fractions can then be rapidly isolated by ultracentrifugation and solubilized using, for example, 1% NP40 or CHAPS. Using this method to identify individual membrane proteins, we have found it useful to employ an ELISA assay to track membrane protein purification, using an antibody to the endogenous tag or marker protein (e.g., succinate dehydrogenase complex) in a 96-well format with homogenates from cells that give a positive response for relative quantification purposes [Woodbury et al., 2002]. scFv will be used to capture the membrane protein and an alkaline phosphatase conjugated polyclonal antibody permits facile detection. The ratio of the relative amounts of the specific membrane protein to total proteins (measured by the Bradford assay) provides an assessment of membrane purity and allows us to follow the progress of the purification process. A method to fractionate the crude membrane preparation would be expected to decrease the complexity of the sample, thus improving the number of proteins identified by each MS run. For this we will subject the crude membrane preparation to IEF using the Rotofor apparatus (BioRad) (pH 3-10 ampholite) run in the presence of 6 M urea. After the focusing reaches equilibrium, 20 fractions will be collected and the amount of protein and pH in each fraction determined. Based on prior experience we expect protein recovery from the fractions to be quantitative with minimal loss to surfaces. Typical total protein yields were 1 mg for the crude cell lysate and 15 ug for the crude membrane protein per gram of packed cells.

2.4.3.2 Affinity Purification of Protein Complexes

Two general affinity approaches will be emphasized to isolate protein complexes from cellular homogenates using mild conditions that are not expected to disrupt protein complexes. Isolation of complexes from proteins

synthesized in vivo and tagged with GST will be done with affinity chromatography on glutathione. These complexes can be eluted with soluble glutathione under mild conditions. In another approach, scFv antibodies to native epitopes will be covalently bound to a stationary phase through a disulfide linkage, permitting capture and release of bound complexes using physiological buffers and mild reducing conditions. Alternatively, affinity resins conjugated to 4',5'-bis (1,2,3-dithioarsolan-2-yl)5-(3-aminopropanamide) fluorescein (i.e., beta-alanyl FlAsH) will be used that recognize unique epitope tags (CCXXCC) that can be engineered either at the N- or C-terminal or within expressed protein sequences [Thorn et al., 2000]. In both of these cases complexes are released under mild reducing conditions (e.g., 50 mM DTT), which will maintain complex formation, and permits a reassessment of the specificity of the complex using a complementary method that may involve either cycling the complex back onto the stationary phase resin or the use of an alternative entry point into the complex. Furthermore, it is expected that the extent of nonspecific protein contamination will be less due to the exquisite specificity and purity of scFv (which does not have a preimmune fraction and can be selectively engineered to reduce nonspecific associations) or because of the specificity of the chemistry associated with the association of CCXXCC with FlAsH. In the latter case FlAsH can be covalently bound to the eluted complex (or following expression of the engineered protein inside *Shewanella*), permitting validation and localization of the protein within a complex under both *in-vitro* and *in vivo* conditions in conjunction with the **Molecular and Cellular Imaging subproject** [Griffin et al., 1999; Sun and Squier, 2000]. These methods will be compared with other reported affinity tags (i.e., GST, FLASH, or the TAP affinity methods described by Gavin and colleagues) that have been used to isolate protein complexes [Smith and Johnson, 1988; Gavin et al., 2002].

scFv Affinity Columns: Standard procedures involving immunoaffinity binding, trypsin-digestion, and MS will be used to isolate and identify protein complexes [Zangar et al., 2002; Smith et al., 2002]. Briefly, antibodies will be covalently attached to a stationary phase using stable or cleavable (disulfide) linkages. Initial studies indicate that Sepharose or silica beads provide good coupling efficiencies and optimal physical characteristics suitable for high-throughput methods. Alternatively, a cross-linked bis-acrylamide/asolactone copolymer (Ultralink™) from Pierce provides a good matrix that in comparison to agarose minimizes nonspecific protein associations [Zangar, 2002]. Stationary phase will be loaded into minicolumns and exposed to concentrated cellular fractions to permit saturation of the binding resin using optical detection to monitor binding (see **High Throughput Complex Processing subproject**).

These methods are in contrast to conventional antibody affinity methods, which require that the antigen complex be eluted with harsh conditions (e.g., mildly acidic solutions of formic acid; pH 2), which is expect to disrupt complex formation. While this latter methodology will serve as a benchmark with respect to the efficiency of standard methods, it is expected that we will emphasize mild conditions involving the reduction of disulfide bonds to release antibody and isolated complex from the column

FlAsH Affinity Purification: The highly specific fluorescent reagent FlAsH (Fluorescein Arsenical Helix binder) has been shown to specifically interact with tetracysteine containing helices in proteins (i.e., CCXXCC) expressed in cells to become a highly fluorescent indicator [Griffin et al., 1998]. Since the binding sequence is rare in natural proteins, exquisite specificity permits both *in vivo* labeling of proteins in cells and the purification of poorly expressed proteins using affinity chromatography [Thorn et al., 2000; Griffin et al., 2000]. Furthermore, the fluorescence signature provides a ready measurement of protein binding that is suitable to detect complex formation of low-abundance proteins in the background of a concentrated cellular fraction. Briefly, FlAsh was modified to contain a primary amine suitable for immobilization onto a stationary phase, permitting rapid purification of expressed protein containing the sequence CCXXCC, which can be located at either the N- or C-terminal or within the expressed protein. Covalent modification is reversible, and captured protein is quantitatively eluted from addition of either 1,2-ethanedithiol (EDT) or dithiothreitol (DTT). The mild elution conditions are "unlikely to perturb protein structure or protein-protein interactions" [Thorn et al., 2000].

GST Affinity Tags: GST polyproteins from *S. oneidensis* or *R. palustris* will be isolated with their associated complex proteins from cell lysates. Cells from one liter cultures will be harvested in log phase growth and induced with IPTG. After induced expression of the plasmid information, whole cells or spheroplasts will be collected. Complexes will be isolated from the appropriate soluble crude cell lysates (periplasm, spheroplast soluble proteins or detergent solubilized membrane proteins) via binding with glutathione agarose beads batchwise at 4°C for 1-2 h. The beads will be harvested and washed at least 4 X 10 ml PBS. The complex will be eluted with 100uL free glutathione at 15mM in 50mM Tris pH8.0/ 10 % glycerol/2mM DTT. Elution is done with rocking at 4°C for 10 min, The eluted protein can be stored frozen at -80°C.

2.4.3.3 Trapping of Complexes Using Intermolecular Cross-linking

Stabilization of protein complexes using intermolecular cross-linking may be necessary to identify weakly associating binding partners, and allows the use of harsher conditions in subsequent purification procedures. Two types of cross-linking experiments will be done: cross-linking of purified complexes and cross-linking of crude cytosolic fractions.

High-Affinity Complexes: Individual protein components within a tightly bound protein complex will be stabilized by cross-linking under dilute conditions (1 mg/mL of protein complex), which will favor the dissociation of nonspecific and weakly association proteins. Cross-linkers will contain a biotin molecule coupled to the spacer arm between the two cross-linking groups. The extent of reaction will be evaluated by using SDS-PAGE. Complexes can be re-purified using the biotin tag or digested directly with trypsin. Following proteolytic digestion, peptides will be identified using MS as described below in Specific Aim 4 and in the **High Throughput Molecular Characterization** and **Bioinformatics and Computing subprojects** [Rappsilber et al., 2000]. Cross-linked peptides can be affinity purified using the biotin tag and a commercially available (Pierce) monomeric avidin affinity column [Gygi et al., 1999; Larsson et al., 2000].

Low-Affinity Complexes: Weakly bound protein associations will be stabilized by carrying out the cross-linking reaction using concentrated (> 20 mg/mL) cytosolic fractions. The lysates can be from either wild-type bacteria or from bacteria containing a plasmid coding for a fusion protein containing a molecular affinity tag. The cross-linked complexes can be purified using either single- or tandem-affinity methods using, for example, the GST or biotin tag. Following complex isolation and proteolytic digestion, the peptides will be identified using mass spectrometry. These data could identify protein associations which are functional, but that occur at low affinity and are difficult to identify by use of other capture techniques.

2.4.4 *Specific Aim 4*. Identify Proteins and Their Stoichiometries in Isolated Complexes.

This aim proposes to identify the proteins, major modifications, and their relative stoichiometries within isolated complexes. This will be accomplished by combining the information from measurements of protein complexes at the intact protein level (i.e. before digestion) with the information from proteolytically processed complexes (identified, for example, using AMTs) [see Appendix; Lee et. al., 2002]. These results will be compared with the results of alternative approaches based upon capillary LC-MS/MS and analyses involving the use of gel electrophoresis separations.

2.4.4.1 Protein Identification and Stoichiometry Using Gel Electrophoresis-based Methods

The most traditional approach to identify individual proteins in an isolated complex involves SDS-PAGE based separation of protein subunits, and excision of protein-containing bands [Anderson and Anderson, 1978a, 1978b; Gianazza et al, 1983]. The proteins in the gel slices are digested with trypsin and analyzed by matrix assisted laser desorption ionization (MALDI) or MS/MS. Since this approach is relatively low-throughput, we expect that the major advantage of gel-based methods will be in the initial estimation of the stoichiometry of protein complexes, as the relative staining intensities of individual bands (correcting for differences in molecular mass) correlates linearly with protein abundance. These results can be used to help rule out nonspecific binding partners that co-purify with tagged-proteins, by use of appropriate controls for non-specific interactions. These gel-based methods will be used to help validate the reliability of the high throughput MS and AMT methods for determining the stoichiometries of proteins in a given complex.

2.4.4.2 High-Throughput Protein Identification Using MS

Separation methods will be coupled with MS to achieve rapid identification of the protein constituents within isolated complexes. The technical approaches under development for MS-based analysis of protein complexes are described in detail in the **High Throughput Molecular Characterization subproject.** A general description of the two approaches currently in use at ORNL and PNNL is presented here to provide a context for the overall process of protein identification in complexes. The relative efficiencies and reproducibility of these general approaches will be compared over the course of the initial characterizations of protein complexes. These comparisons will permit the selection of the optimal methods for high-throughput analysis, which will become

standardized at both laboratories to permit the rapid identification of greater than 80% of all protein complexes in an individual microorganism in timeframes considerably less than one year**.**

The MS identification of the components in protein complexes depends on the capability of software tools to analyze MS/MS results in combination with available genomic information. There are currently a number of such tools commercially available that we routinely use, such as SEQUEST and MASCOT, and we have also developed our own peptide identification software for use with high quality FTICR multiplexed-MS/MS data. Our data management system allows the use of virtually any peptide identification tool, the substitution of one for another at any point, or the combined use of multiple tools for analysis of the same MS/MS data sets. These efforts will use the management system for processing and handling data, as described in Specific Aim #5 and in the **Bioinformatics and Computing subproject**.

Protein Identification Using Accurate Mass Tags: Peptide AMTs is one approach that will be evaluated to provide high throughput and sensitivity for protein mixture characterization [Smith et al., 2001; Smith et al., 2002]. Implementation of this approach to protein identification for complexes purified from *Shewanella* and *Rhodopseudomonas* benefits substantially from the on-going generation of AMT databases for these organisms. Once these libraries are generated, several thousand proteins with large differences in relative abundances can be identified in organisms in a 3 to 10 hour MS analysis (dictated by the time for the capillary LC separation) [Smith et al, 2002; Lipton et al, 2002]. It is expected that protein identification in purified complexes, which will contain a relatively small number of different proteins, will permit significant increases in the speed of the separations used for the measurement of complexes. In modifying our proteomic approach for the characterization of protein complexes we will develop methods that not only provide much faster separations, but will do so using conditions allowing the unchanged use of separation [Smith et al, 2002], as described in the **High Throughput Molecular Characterization subproject**.

Multi-dimensional LC-MS: Two alternative approaches to AMT-based protein identification, which represent current methods that are commonly used for proteomics work, are being assessed with respect to their applicability to the high-throughput identification of protein in isolated complexes. The first approach is anion exchange FPLC fractionation followed by mass spectrometric analysis of each fraction, either at the peptide-level or at the protein level. A crude cell lysate will be fractionated using a quaternary ammonium strong anion exchange column and a linear salt gradient. Fractions (1 mL) will be automatically collected for analysis and stored at –80 °C until needed. A portion of each fraction will be examined by high-resolution ES-FTMS for determination of the molecular masses of the intact proteins. A remaining portion of each fraction will be digested with trypsin using manufacturer's protocols followed by desalting on reverse-phase cartridges and reverse-phase LC-MS/MS for protein identification. The second approach is a two-dimensional LC/LC-MS/MS technique termed "MudPIT" developed by Yates and co-workers [Washburn et al., 2001; Wolters et al., 2001]. Two chromatographic columns with different separation properties are connected serially and interfaced with the electrospray inlet of the Finnigan LCQ instrument. The first stage of separation is based on strong cation exchange chromatography, with the second separation stage involving reverse phase C18 separation. A mobile phase with successively higher salt concentration is applied to elute peptide "plugs" from the cation exchange column, which can then be separated under higher resolution conditions by the downstream reverse-phase column. Our approach is to combine commercially available columns to provide the most robust analytical system, admittedly at the expense of sensitivity. Mass spectra and tandem mass spectra of the eluting peptides are then recorded across the multiple-hour chromatographic run. Post-run processing of the large data sets is performed using SEQUEST and DTASelect to identify the protein from which each peptide originates, and to compile "hits" on a given single protein obtained through different peptides. This approach is able to resolve up to ten thousand peptides in a single run, often providing identification of 1000-2000 corresponding proteins. This technique avoids many of the known difficulties of the gel-based methods, such as limited pH range, run to run irreproducibility, limited dynamic range, solubility limitations for membrane proteins, and limited gel loading capacities, not to mention the vast increase in speed over cutting spots from gels for in-gel digestion. We have already demonstrated this multi-stage LC-MS technique at ORNL in the characterization of the proteome of the bacterium *S. oneidensis* [VerBerkmoes et. al., 2002; see Appendix of **High Throughput Molecular Characterization subproject.**].

2.4.4.3 Determination of Protein Stoichiometries in Isolated Complexes

We will use either capillary LC and/or capillary isoelectric focusing (CIEF) in conjunction with FTICR MS for measurements at the intact protein level [Jensen et al., 1999; Jensen et al., 2000]. CIEF is an attractive separation technique for proteomics due to its speed, efficiency, potential for automation, and high sensitivity when combined with MS. CIEF employs a polyampholyte mixture in free solution that creates a pH gradient in the capillary when an electric field is applied. Proteins migrate through this field until they have zero net charge and

proteins are focused at their respective isoelectric points (pI). After focusing, protein bands are mobilized to the detector using chemical mobilization (anodic or cathodic) or hydraulic mobilization (i.e., pressure, vacuum, or gravity driven flow). Efforts at PNNL have previously shown that >900 putative *E. coli* proteins could be detected in a single low resolution CIEF separation combined with FTICR.

Our approach will be to combine information from protein complex analyses obtained at both the peptide and intact-protein levels. A portion of each complex sample will be digested with trypsin and analyzed as described above to identify peptides and their parent proteins, and at the intact protein level using either CIEF-FTICR or LC-FTICR. In this effort, we will increase the speed of CIEF separations and also explore the relative utility of capillary LC approaches. The peptide analyses will allow us to identify the set of proteins in each complex. In recent work, we have shown that determination of the finite set of proteins in each fraction is sufficient to infer the intact protein's mass [Lee et al., 2002; see Appendix]. In the initial demonstration of this concept we analyzed the yeast large ribosomal complex, and found that the high mass measurement accuracies obtainable from FTICR combined with the identities of the contributing proteins (from analyses at the peptide level) allowed the assignment of nearly all detected species (42 out of 43 of the subunit proteins, and 58 of 64 possible protein isoforms). This approach will provide quantitative information on the abundances of proteins. The efforts under this Task will determine the throughput achievable for the intact protein level analyses, and the facility with which additional information can be gleaned from such analyses. These efforts will determine the utility of intact-protein level analyses in the context of the desired high throughput capability.

2.4.5 *Specific Aim 5*. Integrate Data Management and Protein Complex Analysis

This aim proposes to develop analytical capabilities to permit the accurate identification of protein complexes and verification of their cellular relevance. To accomplish this goal, we will provide transparent linkages between experimental data and developing databases, in conjunction with the **Bioinformatics and Computing subproject**. It will be necessary to implement existing methods and to develop new approaches that permit efficient information retrieval and comparisons between data collected using different experimental methods. It is critical that this include: (1) a sample tracking system to catalog the reagents that are available and the status of different samples within the system, (2) a laboratory information management system (LIMS) that will include information regarding linkages between MS data, growth conditions, and other collected data sets and (3) databases that summarize identified protein complexes and their relationship to existing functional and structural information.

2.4.5.1 Sample Tracking System

Information regarding the status of expressed and tagged proteins will be stored in a regional database that permits facile communication between research teams. Linkages between identified proteins and existing information will be engineered that facilitate data analysis to permit researchers to identify functional linkages between complexes and other cellular systems. Web accessible forms will be used to enter information about the samples as they are processed through the Center. The entry forms will include required fields and pick lists to ensure complete and consistent data collection. The exact processing steps to be tracked and details of data to be collected will be determined in data design sessions that will be held early in the project. Commercially available tracking systems, such as those available from Cimarron (http://www.cimsoft.com), will be evaluated and purchased. To the extent that the commercial systems do not meet the Center's needs, tracking components will be developed internally.

The sample tracking process begins when a researcher defines an experiment, the research goals it supports, the samples to be prepared, and the analyses to be completed for each sample. Following sample preparation, the investigator enters details about the sample including a unique sample identifier, and descriptive information about the sample composition and processing. The analytical lab updates the sample history with confirmation of analytical methods applied to the sample, analysis date, and electronic data files associated with the sample. The sample tracking will be seamlessly interfaced with the Center's Protein Complex Database, where data will be stored following completion of the analytical tasks.

2.4.5.2 Laboratory Information Management System

An open environment that facilitates the identification of linkages between different data sets obtained, for example, under different growth conditions will facilitate data retrieval and pattern recognition that will support high-throughput methodologies. Summarized data will be seamlessly connected to independent databases that summarize expected linkages between different cellular components in a defined context (e.g., growth on benzoate). Data about experimental conditions and various related pieces of information for each run of the mass spectrometer and the raw and processed data will be maintained by software like SEQUEST. Several commercial Laboratory Information Management Systems (LIMS) such as ALIS-2000 by Innovative Automation (http://www.ia-one.com) CimBiosis Proteomics Workflow System by Cimarron (http://www.cimsoft.com), and SQL*LIMS by Applied Biosystems (http://www.appliedbiosystems.com) will be surveyed, evaluated, purchased, and installed in Center labs to facilitate data management. Some customization of the LIMS will be required. For example, data exchange between the Sample Tracking database and the LIMS will need to be implemented. To maintain and allow easy access to the large quantity of raw MS data, each MS laboratory will need to install high-speed filer systems. Based on our rough estimates, one 100 GB to 1 TB filer system for each MS facility will be needed as it enters the production phase, in order to keep the raw data for at least 6 months. Some of the processed data will be transferred to the Center's Protein Complex Database. Early in the project, requirements for data exports to the Protein Complex Database will be established to ensure consistent data reporting across the various instruments and procedures. Adherence to the adopted export specifications is essential for proper data processing and elimination of costly manual corrections.

2.4.5.3 Protein Complex Database

Understanding the context of an isolated protein complex will be critical to our goal of understanding the functional linkages of protein complexes to cell function. It will be important to develop a database that will contain links to all the experimental and computational data analysis, and prediction/modeling results generated in this project. This includes (a) identified proteins, (b) identified PTMs and mutations of each protein, (c) identified protein-protein interactions, (d) interaction maps of protein complexes, (e) complex structures, in low- and high-resolutions, and biological functions, at different levels, and (f) links to related protein databases. The same identification code will be for each protein and each complex among all our databases. The initial data model for this database will be based on existing well-defined data models such as BIND. The database will be implemented as an Oracle database with applications developed using a combination of Perl, PHP, and Oracle 9i Application Server or SQLServer databases.

A number of search and analysis capabilities will be built on top of this Protein Complex Database to facilitate information extraction from outside users. These search tools will be implemented in the early phase of the project and will be made accessible through the Web Interface of the Center. Interfaces will be provided to enable authorized users to extract data from the database, process it through the ORNL toolkit or the PNNL Problem Solving Environment, and optionally store the results of the analysis back into the Protein Complex Database. Additionally, applications will be developed for Protein Complex Database data to be submitted to or pooled with related data in community databases like BIND.

Scientists mining the Center's heterogeneous data sets will require simplified access from many different computing platforms. This access will be provided through a community data portal that will integrate and deliver data and information management, analysis, and sharing capabilities through a web-based interface accessible via a standard web browser. Through the use of profiles, the portal page presentations are customizable to address the specific needs of individuals, research groups, or the broader research community. Basic security and access controls are provided through the portal to ensure that users have the appropriate authorization to access the data and tools. The community portal will include standardized reports as well as *ad hoc* querying capabilities. A significant part of the Center's data will ultimately be made available to the research community. This data will be integrated with other data sources to provide information about related proteins and the genes that express those proteins.

2.5 Subcontract or Consortium Arrangements: None.

2.6 Literature Cited

Anderson, N. G. and Anderson, N. L., "Analytical Techniques for Cell Fractions. XXI. Two-Dimensional Analysis of Serum and Tissue Proteins: Multiple Isoelectric Focusing", Anal. Biochem., 85, 331 (1978).

Anderson, N.L. and Anderson, N.G. , "Analytical Techniques for Cell Fractions. XXI. Two-Dimensional Analysis of Serum and Tissue Proteins: Multiple Gradient Slab-Gel Electrophoresis", Anal. Biochem., 85, 341 (1978).

Anderson, N.L., Nance, S.L., Tollaksen, S.L., Giere, F.A., and Anderson, N.A., "Quantitative Reproducibility of Measurements from Coomassie Blue-stained Two-dimensional Gels: Analysis of Mouse Liver Protein Patterns and a Comparison of BALB/c and C57 Strains", Electrophor., 6, 592 (1985).

Bader, G.D., Donaldson, I., Wolting, C., Ouellette, B.F., Pawson, T., and Hogue, C.W., "BIND--The Biomolecular Interaction Network Database", Nucleic Acids Res., 29, 242 (2001).

Baron, S.F., and Ferry, J.G. , "Purification and Properties of the Membrane-Associated Coenzyme F_{420}- Reducing Hydrogenase from Methanobacerium formicicum", J. Bacteriol., 171, 3846 (1989).

Beliaev, A., Saffarini, D., McLaughlin J., and Hunnicutt, D., "MtrC, an Outer Membrane Decahaem c Cytochrome Required for Metal Reduction in *Shewanella putrefaciens* MR-1", Molecular Microbiology, 39, 722 (2001).

Beliaev, A.S., Thompson, D.K., Khare, T., Lim, H., Brandt, C.C., Li, G., Murray, A.E., Heidelberg, J.F., Giometti, C.S., Yates, J., Nealson, K.H., Tiedje, J.M., and Zhou, J., "Gene and Protein Expression Profiles of *Shewanella oneidensis* During Anaerobic Growth with Different Electron Acceptors", OMICS, 1, 39 (2002).

Belov, M.E., Anderson, G.A., Angell, N.H., Shen,Y., Tolic, N., Udseth, H.R., Smith, R.D. "Dynamic Range Expansion Applied to Mass Spectrometry Based on Data-Dependent Selective Ion Ejection in Capillary Liquid Chromatography Fourier Transform Ion Cyclotron Resonance for Enhanced proteome Characterization," Analytical Chemistry, 73, 5052 (2001).

Carlozzi, P. and Sacchi, A., "Biomass Production and Studies on Rhodopseudomonas Palustris Grown in an Outdoor, Temperature Controlled, Underwater Tubular Photobioreactor", J. Biotechnol., 88, 239 (2001).

Cohen, M., and Yamasaki, H., "Flavonoid-induced Expression of a Symbiosis-related Gene in the Cyanobacterium Nostoc Punctiforme", J. Bacteriol., 182, 4644 (2000).

Davison, J., "Genetic Tools for Pseudomonads, Rhizobia and other Gram-negative Bacteria", Biotechniques, 32, 386 (2002).

Dischert, W., Vignais, P., and Colbeau, A., "The Synthesis of Rhodobacter Capsulatus HupSL Hydrogenase is Regulated by the Two-component HupT/HupR System", Molecular Microbiology, 34, 995 (1999).

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., and Obradovic, Z., "Intrinsically Disordered Protein", J. Mol. Graph. Model.. 19, 26 (2001a).

Dunker, A.K. and Obradovic, Z., "The Protein Trinity--linking Function and Disorder", Nat. Biotechnol., 19, 805 (2001b).

Dyson, H.J. and Wright, P.E., "Coupling of Folding and Binding for Unstructured Proteins", Curr. Opin.Struct. Biol., 12, 54 (2002).

Einhauer, A. and Jungbauer, A., "The FLAG™ Peptide, a Versatile Fusion Tag for the Purification of Recombinant Proteins", J. Biochem. Biophys. Methods, 49, 455 (2001).

Ellis, R.J., "Macromolecular Crowding: Obvious but Underappreciated", Trends Biochem. Sci., 26, 597 (2001).

Elsen, S., Dischert, W., Colbeau, A., and Bauer, C., "Expression of Uptake Hydrogenase and Molybdenum Nitrogenase in Rhodobacter Capsulatus Is Co-regulated by the RegB-RegA Two-component Regulatory System", J. Bacteriol., 182, 2831 (2000).

Field, S.J., Dobbin, P.S., Cheesman, M.R., Watmough, N.J., Thomson, A.J., and Richardson, D.J., "Purification and Magneto-optical Spectroscopic Characterization of Cytoplasmic Membrane and Outer Membrane Multiheme c-Type Cytochromes from *Shewanella frigidimarina* NCIMB400", J. Biol. Chem., 275, 8515 (2000).

Friedrich, C., Quentmeier, A., Bardischewsky, F., Rother, D., Kraft, R., Kostka, S. and Prinz, H.. "Novel Genes Coding for Lithotrophic Sulfur Oxidation of Paracoccus Pantotrophus GB17", J. Bacteriol., 182, 4677 (2000).

Garner, E., Romero, P., Dunker, A.K., Brown, C., Obradovic, Z., "Predicting Binding Regions within Disordered Proteins", Genome Inform. Ser. Workshop Genome Inform., 10, 41 (1999).

Gaasterland, T. and Ragan, M.A., "Microbial Genescapes: Phyletic and Functional Patterns of ORF Distribution Among Prokaryotes", Microb. Comp. Genomics, 3, 199 (1998).

Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.-A., Copley, R. R., Edelmann, A., Querfurty, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G., "Functional Organization of the Yeast Proteome by Systematic Analysis of Protein Complexes", Nature, 415, 141 (2002).

Gianazza, E., Artoni, G., and Righetti, P.G., "Isoelectric Focusing in Immobilized pH Gradients in Presence of Urea and Neutral Detergents", Electrophor.. 4, 321 (1983).

Griffin, B.A., Adams, S.R., and Tsien, R.Y., "Specific Covalent Labeling of Recombinant Protein Molecules Inside Living Cells", Science, 281, 269 (1999).

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., "Quantitative Analysis of Complex Protein Mixtures Using Isotope-coded Affinity Tags", Nature Biotech., 17, 994 (1999).

Hemmingsen, S., Woolford, C., van der Vies, S., Tilly, K., Dennis, D., Georgopoulos, C., Hendrix, R., and Ellis, R., "Homologous Plant and Bacterial Proteins Chaperone Oligomeric Protein Assembly", Nature, 333, 330 (1988).

Ho, Y., Gruhler, A., Heilbut. A., Bader, G.D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo. M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., lin, Z., Michalickova, K., Wwilliams, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Pudtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W.V., Figeys, D., and Tyers, M., "Systematic Identification of Protein Complexes in *Saccharomyces cerevisiae* by Mass Spectrometry", Nature, 415, 180 (2002).

Iakoucheva, L.M., Kimzey, A.L., Masselon, C.D., Bruce, J.E., Garner, E.C., Brown, C.J., Dunker, A.K., Smith, R.D., and Ackerman, E.J., "Identification of Intrinsic Order and Disorder in the DNA Repair Protein XPA", Protein Sci., 10, 560 (2001).

Ikegami, A. Nakasone, K., Kato, C., Nakamura, Y., Yoshikawa, I., Usami, R., Horikoshi, K., "Glutamine Synthetase Gene Expression at Elevated Hydrostatic Pressure in a Deep-sea Piezophilic *Shewanella violacea*",FEMES Microbiol. Lett., 192, 91 (2000).

Inui, M., Roh, J., Zahn, K., Yukawa, H., "Sequence Analysis of the Cryptic Plasmid pMG101 from *Rhodopseudomonas palustris* and Construction of Stable Cloning Vectors", Appl. Environ. Microbiol., 66, 54 (2000).

Jensen, P.K., Pasa-Tolic, L., Anderson, G.A., Horner, J.A., Lipton, M.S., Bruce, J.E., Smith R.D., "Probing Proteomes using Capillary Isoelectric Focusing-electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", Anal. Chem., 71, 2076 (1999).

Jensen, P.K., Pasa-Tolic, L., Peden, K.K., Martinovic, S., Lipton, M.S., Anderson, G.A., Tolic, N., Wong, K.K., and Smith, R.D., "Mass Spectrometric Detection for Capillary Isoelectric Focusing Separations of Complex Protein Mixtures", Electrophoresis, 21, 1372 (2000).

Johansen, C., Verheul, A., Gram, L., Gill, T., and Abee, T., "Protamine-induced Permiabilization of Cell Envelopes of Gram-positive and Gram-negative Bacteria", Appl. Environ. Microbiol., 63, 1155 (1997).

Joshi, H.M. and Tabita, F. R., "A Global Two Component Signal Transduction System that Integrates the Control of Photosynthesis, Carbon Dioxide Assimilation, and Nitrogen Fixation", Proc. Natl. Acad. Sci., 93, 14515 (1996).

Kelleher, N.L., Lin, H.Y., Valaskovic, G.A., Aaserud, D.J., Fridriksson, E.K., and McLafferty, F.W., "Top Down Versus Bottom up Protein Characterization by Tandem High-resolution Mass Spectrometry", J. Am. Chem. Soc., 121, 806 (1999).

Knablein, J., Dobbek, H. and Schneider, F., "Organization of the DMSO Respiratory Operon of *Rhodobacter capsulatus* and its Consequences for Homologous Expression of DMSOR/TMAOR", Biol. Chem., 378, 303 (1997).

Kocarek, T.A., Zangar, R.C., Novak, R.F., "Post-transcriptional Regulation of Rat CYP2E1 Expression: Role of CYP2E1 mRNA Untranslated Regions in Control of Translational Efficiency and Message Stability", Arch. Biochem. Biophys., 376, 180 (2000).

Krause, B. and Nealson, K.H., "Physiology and Enzymology Involved in Denitrification by *Shewanella putrefaciens*", Appl. Environ. Microbiol., 63, 2613 (1997).

Kumar, A. and Snyder, M., "Protein Complexes Take the Bait", Nature, 415, 123 (2002).

Lancaster, C.R., "Succinate:Quinone Oxidoreductases: An Overview", Biochim. Biophys. Acta,. 1553, 1 (2002).

Larsson, T., Bergström, J., Nilsson, C., Karlsson, K.-A., "Use of an Affinity Proteomics Approach for the Identification of Low-abundant Bacterial Adhesins as Applied on the Lewis^{b}-binding adhesin of *Helicobacter pylori*",FEBS Lett., 469, 155 (2000).

Lathe, W.C., 3rd, Snel, B., and Bork, P., "Gene Context Conservation of a Higher Order than Operons", Trends Biochem. Sci., 25, 474 (2000).

Lee, S.-W., Berger, S.J., Marinovic, S., Pasa-Tolic, L., Anderson, G.A., Shen, Y., Zhao, R., and Smith, R. D., "Direct Mass Spectrometric Analysis of Intact Proteins of the Yeast Large Ribosomal Subunit Using Capillary LC/FTIRC", Proc. Natl. Acad. Sci. U.S.A., in press (2002).

Leys, D., Tsapin, A.S., Nealson, K.H., Meyer, T.E., Cusanovich, M.A., Van Beeumen, J.J., "Structure and Mechanism of the Flavocytochrome c Fumarate Reductase of *Shewanella putrefaciens* MR-1", Nat. Struct. Biol., 6, 1113 (1999).

Liou, G.G., Jane, W.N., Cohen, S.N., Lin, N.S., and Lin-Chao, S., "RNA Degradosomes Exist *in vivo* in Escherichia coli as Multicomponent Complexes Associated with the Cytoplasmic Membrane via the N-terminal Region of Ribonuclease E", Proc.Natl. Acad .Sci. U S A, 98, 63 (2001).

Lipton, M.S., Pasa-Tolic, L., Anerson, G.A., Anderson, D.J., Auberry, D.L., Battista, J.R., Daly, M.J., Fredrickson, J., Hixson, K.K., Kostandarithes, H., Masselon, C.D., Markille, L.M., Morre, R.J., Romine, M.F., Shen, Y., Tolic, N., Udseth, H.R., Venkateswaran, A., Wong, K.K., Zhao, R., and Smith, R. D., "Global Analysis of *Deinococcus radiodurans* R1 Proteome Using Accurate Mass Tags", Proc. Natl. Acad. Sci. U.S.A., submitted (2002)

Liu, Q., Li, M.Z., Leibham, D., Cortez, D., and Elledge, S.J., "The Univector Plasmid-fusion System, A Method for Rapid Construction of Recombinant DNA Without Restriction Enzymes", Curr. Biol., 8, 1300 (1998).

Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O., Eisenberg, D., "Detecting Protein Function and Protein-protein Interactions from Genome Sequences", Science, 285, 751 (1999).

Myers, C.R. and Myers, J.M., "Localization of Cytochromes to the Outer Membrane of Anaerobically Grown *Shewanella putrefaciens* MR-1", J. Bacteriol., 174, 3429 (1992).

Myers, J.M., and Myers, C.R., "Role of the Tetraheme Cytochrome CymA in Anaerobic Electron Transport in Cells of *Shewanella purtrfaciens* MR-1 with Normal Levels of Menaquinone", J. Bacteriol., 182, 67 (2000).

Newman, D.K. and Kolter, R., "A Role for Excreted Quinones in Extracellular Electron Transfer", Nature, 405, 94 (2000).

O'Farrell, P.H., "High-resolution Two-dimensional Electrophoresis of Proteins", J. Biol. Chem., 250, 4007 (1975).

Pellegrini, M., Marcotte, E.M., and Yeates, T.O., "A Fast Algorithm for Genome-wide Analysis of Proteins with Repeated Sequences", Proteins, 35, 440 (1999).

Pelletier, D.A. and Harwood, C.S., "2-Hydroxycyclohexanecarboxyl A Dehydrogenase, an Enzyme Characteristic of the Anaeorbic Benzoate Degradation Pathway Used by *Rhodopseudomonas palustis*", J. Bacteriol., 182, 2753 (2000).

Pratt, J.M.. Transcription and Translation. IRL Press (1984).

Puig, O., Caspary, F., Rigaut, G. Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M, and Seraphin, B., "The Tandem Affintity Purification (TAP) Method: A General Procedure of Protein Complex Purification", Methods, 24, 218 (2001).

Qian, Y. and Tabita, F., "Expression of glnB and glnB-like Gene (glnK) in a Ribulose Bisphosphate Carboxylase/oxygenase-deficient Mutant of *Rhodobacter sphaeroides*", J. Bacteriol., 180, 4644 (1998).

Ramagli, L.S. and Rodriguez, L.V., "Quantitation of Microgram Amounts of Protein in Two-dimensional Polyacrylamide Gel Electrophoresis Sample Buffer", Electrophor., 6, 559 (1985).

Rappsilber, J., Siniossoglou, S., Hurt, E., and Mann, M.A., "Generic Strategy to Analyze the Spatial Organization of Multi-protein Complexes by Cross-linking and Masss Spectrometry", Anal. Chem., 72, 267 (2000).

Reid, G.A., Miles, C.S., Moysey, R.K., Pankhurst, K.L., and Chapman, S.K., "Catalysis in Fumarate Reductase".Biochim. Biophys. Acta, 1459, 310 (2000).

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B., "A Generic Protein Purification Method for Protein Complex Characterization and Proteome Exploration", Nature Biotechnology, 17, 1030 (1999).

Séraphin, B., Puig, O., Bouveret, E., Rutz. B., and Caspary, F., Tandem Affinity Purification to Enhance Interacting Protein Identification in Protein-Protein Interactions, (E. Golemis, Ed.) pp. 313-328. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2002).

Smith, D.B. and Johnson, K.S., "Single-step Purification of Polypeptides Expressed in Escherichia coli as Fusions with Glutathione S-transferase", Gene, 67, 31 (1988).

Smith, R.D., Anderson, G.A., Lipton, M.S., Masselon, C., Pasa-Tolic, L., Shen, Y., and Udseth, H.R., "The Use of Accurate Mass Tags for High-throughput Microbial Proteomics", OMICS, 6, 61 (2002).

Smith, R.D., Pasa-Tolic, L., Lipton, M.S., Jensen, P.K., Anderson, G.A., Shen, Y., Conrads, T.P., Udseth, H.R., Harkewicz, R., Belov, M.E., Masselon, C., and Veenstra, T.D., "Rapid Quantitative Measurements of Proteomes by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry",Electrophoresis, 22, 1652 (2001).

Sun, H. and Squier, T.C., "Ordered and Cooperative Binding of Opposing Globular Domains of Calmodulin to the Plasma Membrane Ca-ATPase", J. Biol. Chem., 275,1731 (2000).

Thompson, D.K., Beliaev, S., Giometti, C.S., Tollaksen, S.L., Khare, T., Lies, D.P., Nealson, K.H., Lim, H., Yates, J., Brandt, C.C., Tiedje, J.M., and Zhou, J., "Transcriptional and Proteomic Analysis of a Ferric Uptake Regulator (Fur) Mutant of *Shewanella oneidensis*: Possible Involvement of Fur in Energy Metabolism, Transcriptional Regulation, and Oxidative Stress", Appl. Environ. Microbiol., 68, 881 (2002).

Thorn, K.S., Naber, N., Mtuska, M., Vale, R.D., and Cooke, R., "A Novel Method of Affinity-Purifying Proteins Using a bis-arsenical Fluorescein", Prot. Sci., 9, 213 (2000).

Tichi, M.A., Meijer, W.G., and Tabita, F.R., "Complex I and its Involvement in Redox Homeostasis and Carbon and Nitrogen Metabolism in *Rhodobacter capsulatus*", J. Bacteriol., 183, 7285 (2001).

Tsapin, A., Vandenberghe, I., Nealson, K., Scott, J., Meyer, T., Cusanovich, M., Harada, E., Kaizu, T., Akutsu, H., Leys, D., Van Beeumen, J., "Identification of a Small Tetraheme Cytochrome c and a Flavocytochrome c as Two of the Principal Soluble Cytochromes c in *Shewanella oneidensis* Strain MR1", Appl. Environ. Microbiol., 67, 3236 (2001)..

Verberkmoes, N.C., Bundy, J.L., Hauser,L., Asano, K.G., Razumovskaya, J., Larimer, F., Hettich, R.L., Stephenson, J.L. Jr., "Integrating "Top-down" and "Bottom-up" Mass Spectrometric Approaches for Proteomic Analysis of *Shewanella oneidensis*", J. Proteome Research, in press (2002).

Washburn, M.P., Wolters, D., and Yates, J.R. 3rd, "Large-scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology", Nat. Biotechnol., 19, 242 (2001).

Wolters, D.A., Washburn, M.P., and Yates, J.R. 3rd, "An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics", Anal. Chem., 73, 5683 (2001).

Woodbury, R.L., Varnum, S.M., and Zangar, R.C., "Elevated HGF Levels in Sera from Breast Cancer Patients Detected Using a Protein Microarray ELISA", Molecular Pharmacology, in press (2002).

Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M., and Eisenberg, D., "DIP, the Database of Interacting Proteins: A Research Tool for Studying Cellular Networks of Protein Interactions", Nucleic Acids Res., 30, 303 (2002).

Xu, Y. and Xu, D., "Protein Threading Using PROSPECT: Design and Evaluation", Proteins 40, 343 (2000).

Xu, Y, Xu, D. and Olman, V., "A Practical Method for Interpretation of Threading Scores: An Application of Neural Networks", Statistica Sinica Special issue on Bioinformatics, in press (2002).

Yang, T., Horejsh, D.R., Mahan, K.J., Zaluzec, E.J., Watson, J.T., and Gage, D.A., "Mapping Cross-linking Sites in Modified Proteins with Mass Spectrometry: An Application to Cross-linked Hemoglobins", Anal. Biochem., 242, 55 (1996).

Zangar, R.C., Kimzey, A.L., Okita, J.R., Wunschel, D.S., Edwards, R.J., Kim, H., and Okita, R.T., "Cytochrome P450 3A Conjugation to Ubiquitin in a Process Distinct from Classical Ubiquitination Pathway", Mol. Pharmacol., 61, 892 (2002).

2.8 Milestones

Task Summary and Milestones

Tasks	FY03		FY04		FY05	
Initial target clones built	x	x				
Complete clone library		x	x	x		
Express target clones		x	x	x		
Optimize complex purification techniques		x	x	x	x	x
ID protein complexes by MS	x	x	x	x	x	x
Install/optimize sample tracking system, LIMS for mass spectrometry	x	x	x	x		
Protein complex database, BIND			x	x	x	x

FY03 - Establish Center, incorporating existing separation and analysis capabilities, sample tracking, data handling, and data archival

FY04: Assess and refine analytical and computational tools for the Center and complete the characterization and comparison of an initial set of protein complexes in *S. oneidensis* and *R. plaustris*

FY05 – Continue to Assess and refine analytical and computational tools for the Center and continue to identify protein complexes in the two target organisms, with longer term goal to complete organisms at end of five years.