4.0 HIGH THROUGHPUT MOLECULAR CHARACTERIZATION

Abstract

This subproject focuses on alternative or significantly improved approaches for high throughput molecular characterization of protein complexes to address Goal 1 of the DOE's Genomes to Life Program. The subproject is composed of several research activities, each offering the potential for significant gains in the throughput or information content from protein complex identification. This subproject takes advantage of unique resources in biological mass spectrometry at both Oak Ridge National Laboratory (ORNL) and Pacific Northwest National Laboratory (PNNL) and special expertise from Sandia National Laboratory, the University of Utah, and the University of North Carolina. One approach will aim toward improving throughput of protein complex measurement, allowing scalability to levels that will be required by the GTL Program. We propose a method, based on chemical crosslinking and mass spectrometry, that will augment the use of genetically-engineered, affinity-tagged "bait" proteins for isolation of protein complexes from microbes. Alternative methods for protein complex analysis, including FPLC fractionation and simultaneous accurate mass and MS/MS spectral acquisition, are proposed. A second approach focuses on extending the "accurate mass tag" (AMT) technique developed at PNNL by enhancing the measurement sensitivity and dynamic range of this approach. These improvements should both increase throughput and reduce the sample size requirement, which is important for lower-abundance protein complexes. In addition, efforts will also be focused on developing approaches for improved protein identifications, better absolute protein abundance measurements and improved assessments of data quality. The different approaches taken in this subproject will provide opportunities for validation of new mass spectrometric methods for high-throughput protein complex analysis.

Principal Investigators:

ORNL:
Dr. Gregory B. Hurst
Chemical Sciences Division
Oak Ridge National Laboratory
P. O. Box 2008, MS-6365
Oak Ridge, TN 37831-6365
Phone: 865/574-6142
Facsimile: 865/576-8559
E-mail: hurstgb@ornl.gov

PNNL:
Dr. Richard D. Smith
Environmental Molecular Sciences Laboratory
Pacific Northwest National Laboratory
P.O. Box 999, MS K8-98
Richland, WA 99352
Phone: 509/376-0723
Facsimile: 509/376-7722
E-mail: dick.smith@pnl.gov

4.2 Background and Significance

The recent dramatic expansion in the application of mass spectrometry (MS) to proteomics signals great promise for continuing benefits to the biology community. Discovery-based proteomics research has indeed relied on the enabling technology of MS. A challenge set forth in the GTL program is to move beyond a catalog of proteins in an organism, and to begin identifying all protein interactions and complexes. In this proposal, we outline a series of mass spectrometry-based advances that extend the capabilities described in the **Center for Molecular and Cellular Systems Subproject**. Some of these advances will build on demonstrated technologies for characterizing complex mixtures of proteins, while others offer novel approaches to addressing the issue of high throughput for identifying the suite of protein complexes in an organism.

Several approaches have been used for studying protein complexes, including the yeast two-hybrid system [Fields, 1989; Uetz, 2000], phage display [Rodi, 1999], and protein "chips" [de Wildt, 2000; MacBeath, 2000; Zhu, 2000; Reineke, 2001]. Recently, extensive results have been described for isolating protein complexes from cells by expressing one component of the complex as a fusion protein with affinity tags [Rigaut, 1999; Gavin, 2002; Ho, 2002]. The tagged protein, along with co-purifying proteins present in complexes with the tagged component, is then purified from a cell lysate and the components characterized using mass spectrometry. At this time MS appears most likely to be the primary technology used for protein complex characterization. Mass spectrometry has been used e.g. to study intact non-covalent protein complexes [Loo, 1997; Smith, 1997; Miranker, 2000], providing information on the stoichiometry of the complex and the molecular masses of the component proteins [Loo, 2000]. Most protein characterization approaches now employ separation methodologies combined with some form of MS, typically applied after proteolytic digestion. Identification of peptides sufficiently large for protein identification, typically >~5 to 10-mer size, is now effectively achieved by MS. A widely used approach involves MS selection of a peptide that is dissociated to form fragments followed by mass-to-charge ratio (*m/z*) measurements (i.e. MS/MS) [Eng, 1994; Mann, 1994]. Measurements at the peptide level clearly constitute the most effective approach for protein identification and have been the basis for nearly all initial efforts to obtain increased throughput for protein complex characterization.

The characterization of protein complexes is a challenging analytical problem, and it is clear that current capabilities for the characterization of protein complexes leave much to be desired. First, the throughput of protein complex measurements is low, currently corresponding to approximately a day or more per complex using a single mass spectrometer. Second, the data quality is questionable, as is evident from the limited overlap of protein identifications obtained from the first broad applications to yeast protein complexes [Gavin, 2002; Ho, 2002]. Presently, it is not clear how various factors, including protein “pull down” methodologies, contributed to the dramatic differences in protein constituents identified for the common complexes studied. Since isolation and purification techniques for protein complexes will not be perfect, one must design an approach that will allow detection of low-abundance proteins from the “desired” complex in a sea of background proteins. Additionally, the low level peptides and proteins associated with complexes may carry important information (e.g. on weakly associated proteins, and potentially on protein localization). Beyond this, important protein modification states may be present at lower relative abundances. An additional challenge arises from the very limited sample sizes that will be available for protein complexes, particularly as efforts increase throughput. For these reasons multiple measurements, multiple experimental approaches, and bioinformatics techniques will clearly all be required to generate validated information on protein complexes.

The major objective for the efforts described in this subproject is to significantly advance current mass spectrometric technologies for incorporation into the **Center for Molecular and Cellular Systems Subproject** for greater throughput and/or additional data on protein complexes. This subproject is composed of several distinct efforts that each offers the potential for significant gains in the throughput or information content from protein characterization measurements, and which share the common feature of involving mass spectrometry. The work described in this subproject will be driven by the needs of the **Center for Molecular and Cellular Systems Subproject**. Researchers in this subproject are also involved in the core operations of the Center. For example, PNNL is evaluating a strategy for protein complex analysis based on modification and application of a faster version of the combination of instrumental and methodological approaches presently being developed for

proteomics using accurate mass tags. **Center for Molecular and Cellular Systems Subproject** also describes ORNL implementation of "top-down, bottom-up" characterization of protein complexes; this approach relies on anion exchange FPLC fractionation, followed by intact protein and peptide-level mass measurements. Additionally, Prof. Raymond Gesteland (University of Utah) and co-workers, and Prof. Michael Giddings, University of North Carolina, will provide additional methods for measurement of intact protein mass, and integration of these results with other data to increase robustness of protein identification.

As outlined in the Overview to this proposal, the projects described are the initial research projects being proposed to support the Center for Molecular and Cellular Systems. These projects will be reviewed annually as outlined in the Management Overview for progress and continued relevance to the Center. Normally no project will extend beyond two years. It should be noted that techniques developed as part of this project will be incorporated into both Center sites for the high throughput analysis of protein complexes.

4.2.1 Extension of the Accurate Mass Tag Approach for Protein Identification

Efforts at PNNL will be focused in two areas and aim to increase the utility of the "accurate mass tag" (AMT) concept for protein complex characterization, which is one of two approaches initially being evaluated in the **Center for Molecular and Cellular Systems**. These efforts aim to increase the sensitivity and dynamic range of measurements, and serve to both increase the throughput of measurements and reduce the sample size requirements, an important requirement for protein complexes having lower abundances. The efforts will also develop approaches for obtaining improved protein identifications, better absolute protein abundance measurements and improved assessments of data quality. These efforts will address the important area of the confidence for protein complex characterizations and the issues of data quality (e.g. of relative and absolute protein abundances) that are crucial for data use in understanding and modeling aspects of cellular systems.

The achievable sensitivity and dynamic range of measurements will largely define the scope and utility of the analytical approach. High sensitivity is important because many protein complexes with significant roles in cellular processes will be present only at modest copy number. Thus, it is critical that protein complex characterization be sensitive enough for characterization of complexes from reasonable cell populations. Furthermore, the methods to be used should offer sufficient dynamic range to detect and quantify differences in low abundance proteins associated with complexes, provide information on their modification states, and effectively cover cell surface/membrane proteins. In the context of the AMT approach developed at PNNL, an increase in dynamic range allows protein complexes to be characterized more rapidly (since it is less important that low abundance components be first separated from high abundance components). An increase in dynamic range also increases the effective sensitivity for components not resolved using separations, and can potentially also improve the quantitative utility of the data.

Finally, developments will be pursued at PNNL to increase the quality and reliability of protein complex data due to their significance in the use of the data for modeling, simulating and understanding microbial processes at the molecular level. The initial efforts in the Center will use Fourier Transform Ion Cyclotron Resonance (FTICR) mass spectrometry capabilities for the characterization of protein complexes. However, efforts under this subproject will also explore the utility of another instrumental approach using less expensive instrumentation based upon commercially available quadrupole-time of flight (Q-TOF) mass spectrometers that could apply the AMT approach with even greater throughput and reduced cost. The efforts in this subproject will apply artificial neural nets and statistical methods to aid the identification of proteins and also increase the quality and utility of data by e.g. using elution time information from separations more effectively and to extract more quantitative information from mass spectrometric results. These efforts will mesh with the **Bioinformatics and Computing Subproject**, and aims to provide data of higher quality and more statistically defined quality.

4.2.2 Scaling Mass Spectrometry for High Throughput

In proteomics applications of mass spectrometry, revolutionary advances are resulting from making a leap from "one-at-a-time" measurements that involve, for example, cutting individual protein spots from 2D gels, digesting, and MS analysis, to "shotgun" measurements that involve digesting an unseparated mixture of proteins, and then analyzing the resulting peptides via chromatographic separation with MS detection [Washburn, 2001; Wolters, 2001]. An individual peptide eluting from the chromatographic system and analyzed by MS (both single-stage and tandem) can often provide information to identify the protein from which that peptide originated [Eng, 1994; Mann, 1994]. Compiling this information for all peptides gives a picture of the proteome that rivals or surpasses, in both speed and quality, that obtained by the 2D gel approach.

We propose to explore feasibility of a similar "leap" for analyzing protein complexes. The initial approach to be taken in the **Center for Molecular and Cellular Systems Subproject** will be to isolate individual complexes, based on established affinity techniques–an approach that will provide necessary initial Goal 1 successes. However, to progress beyond the single-complex-at-a-time scale, approaches are required for treating multiple, and perhaps many, complexes simultaneously, without using reagents specific to each complex. A combination of affinity-labeled chemical crosslinkers with modern proteomics techniques offers the potential for such an approach.

Additionally, it is important when designing high throughput approaches to consider opportunities for performing operations in parallel. We propose tasks that will explore the use of arrays of protein affinity capture devices interfaced with electrospray mass spectrometry, and would be compatible with arrays of miniature ion trap mass spectrometers, a combination that could potentially perform mass analysis on many samples simultaneously.

Validation of data on protein complexes by comparing results from different methods will be important for the entire GTL program. We will investigate alternative approaches to mass spectrometric characterization of complexes, including preservation of protein complexes through anion exchange fractionation of cell lysates, and split-effluent chromatography combined with simultaneous high-resolution and MS/MS measurements.

4.3 Preliminary Studies

Initial demonstration of the use of AMTs for proteomics. The speed or coverage ultimately obtained by combined separation-MS approaches depends substantially on the achievable dynamic range of the MS measurements and the resolution (or peak capacity) of the separation step(s) preceding the MS. The combined effective resolving power supplied by the combination of capillary LC separations and FTICR allows the level of complexity presented by protein complexes to be rapidly addressed. In the initial demonstration (see Appendix) of our approach for comprehensive proteome studies of *D. radiodurans* the generation of AMTs involved a two-stage process. The proteome was digested (e.g. with trypsin) and analyzed by capillary LC-MS/MS (generally using quadrupole ion traps), and where, during separations, 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 yielded "potential mass tags" (PMTs) that are subsequently validated as AMTs if the predicted peptide's accurate mass is observed using FTICR at an equivalent LC elution time. PMTs were tentatively identified using "scores" from the search program SEQUEST (based upon the similarity of the spectrum with a set of peaks predicted on the basis of the known most common peptide fragmentation processes) [Smith, Anderson et al., 2002]. Such scores range from low scores yielding highly doubtful identifications, to high scores yielding quite reliable identifications, with no clear line of demarcation. Selecting only the highest scores reduces false positives, but fewer proteins are identified. Conventionally, many MS/MS spectra and search results are manually examined to evaluate spectrum quality and ranking of peptide scores, generally resulting in discard of a substantial fraction of peptide identifications. Our approach using highly accurate mass measurements provides an additional and high quality "test" for a tentative peptide, and significantly increases confidence of identifications [Smith, Anderson et al., 2002]. Once a protein has been identified using AMTs, without need to re-establish peptide identifications using MS/MS analyses, high-throughput studies of protein complexes become possible. For example, using SEQUEST only ~70% of peptides with scores above 2.0 were validated as AMTs for *D. radiodurans*. An analysis of the conversion of PMTs to AMTs as a function of SEQUEST quality score for all peptides identified from ion trap MS/MS measurements showed that most PMTs having a SEQUEST cross

correlation (X_{corr}) score >4.0 were converted to AMTs. A rapid decrease in the conversion of PMTs to AMTs, however, is observed for PMTs identified with an X_{corr} value less than 3.0, to <40% for scores of ~2 [Smith, Anderson et al., 2002]. This result illustrates the decreased reliability of peptide identifications for intermediate scores (which often account for the majority of peptides identified). These results demonstrate the significant increase in confidence of peptide identifications obtained through applying additional accurate mass information.

The use of AMTs has been initially demonstrated for the small radiation resistant, prokaryotic organism *Deinococcus radiodurans*. *D. radiodurans* has a ~3.1 Mbase genome that initial annotation efforts *predict* 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 the organism 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 proteome 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.

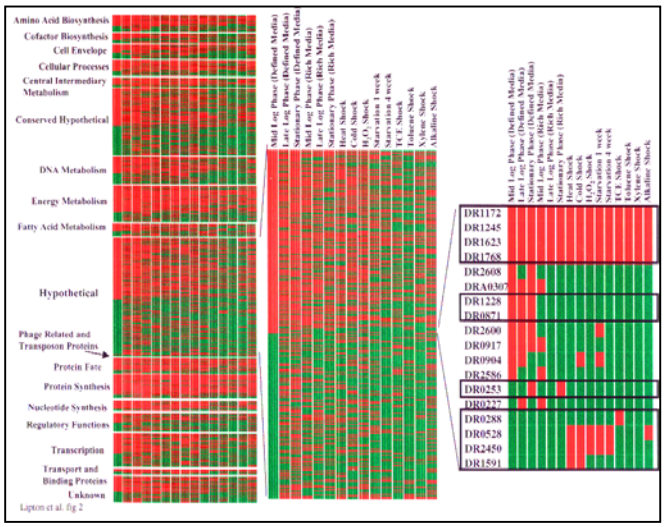

Fig. 4.3.1 Left: The qualitative pattern of ORF expression observed using capillary LC-FTICR by TIGR assigned functional category (red=detected; green=not observed) [White, Eisen et al. 1999]. Each column corresponds to a specific culture condition. **Middle:** An expansion of the predicted "hypothetical" proteins actually detected, illustrating similarities and differences in the patterns for protein expression under different culture conditions (**right**). For a larger version of this figure, please see appendix.

The potential for high-throughput analyses was also illustrated in this initial work. The use of AMTs allowed a global proteomic study of a wide variety (>15) of culture conditions. **Figure 4.3.1** illustrates the qualitative expression pattern for all the detected proteins under the various conditions. Examination of the patterns can suggest potential functions for previously "unknown" proteins. For example, many "hypothetical" proteins were expressed under every culture condition, suggesting a "housekeeping" function (e.g. *D. radiodurans* ORFs DR1172, DR1245, DR1623 and DR1768; **Fig. 4.3.1**, right), while DR0871 and DR1228 were expressed only in minimal media, and DR0528, DR1591 and DR2450 expressed only under certain "stressed" conditions.

The AMT concept can be readily extended to post-translational modifications, sequence variations, frame shifts, etc., using the MS/MS and accurate mass data sets obtained during AMT data base generation. The generation and use of AMTs enables high-throughput and high precision expression studies based on stable-isotope labeling by directly comparing protein complexes e.g., obtained from different culture conditions (utilizing a labeled control or "reference proteome"). Stable-isotope labeled reference proteins from a specified system provide an effective internal standard for each protein, and allow changes in protein abundances based upon the relative abundances of detected species to be assessed. An important aspect of the proposed research will be improving the utility of such data for modeling and simulation by both increasing the quality of the data and providing an improved assessment of data quality. Such developments are critically important for the GTL program since large quantities of data will be generated both for determining the most effective techniques and refining the overall methodology, as well as for its application to multiple microbial systems.

Dynamic range improvements to enhance throughput, coverage and data quality: We propose to build upon our recently developed method for expanding the dynamic range of FTICR measurements to increase the effective

throughput, sensitivity and quality of measurements for protein complexes. We recently reported an approach for dynamic range expansion that involves ion ejection by resonant rf-only dipolar excitation from a linear (2-D) quadrupole device located external to the FTICR ion trap. We selectively remove the major species in this quadrupole so that only the lower abundance species are accumulated, and thus better exploit the dynamic range of FTICR. Initial results show that this Dynamic Range Enhancement Applied to Mass Spectrometry (DREAMS) approach provides a significant gain in the coverage of proteomic measurements. Our initial DREAMS demonstration used mass spectra acquired using standard "non-selective" ion accumulation, in which every second spectrum is preceded by software-controlled removal of the most abundant peaks prior to ion accumulation (to obtain the next spectrum immediately following the non-selective "normal" spectrum). We initially evaluated DREAMS for characterizing a global yeast proteome tryptic digest. We generated two data sets comprising spectra for the non-selective and selective DREAMS accumulations, and compared the results to an analysis using only the standard non-selective ion accumulation method. The number of peptides detected with the alternating sequences (30,771) was about 35% higher than that acquired using the non-selective ion accumulation (22,664). The same methodology provided a 40% increase in the number of peptides using data-dependent ejection of the three most abundant species. Subsequently we implemented DREAMS with high performance capillary LC-FTICR mass spectrometry and examined a global tryptic digest of ^{14}N- and ^{15}N-labeled peptides from mouse B16 cells initially mixed in approximately 1:1 ratio. Two chromatograms can be reconstructed from the normal and DREAMS spectra. **Figure 4.3.2** illustrates the dynamic range enhancement capability using DREAMS for one point in a capillary LC separation where the normal spectrum (**Fig. 4.3.2**, middle top) is dominated by a number of major peptide ions. The information from this spectrum was used "on the fly" to remove these species during ion accumulation for the next spectrum. As a result, the most abundant species were ejected prior to accumulation and the next FTICR spectrum was dominated by a much different set of species (**Fig. 4.3.2**, middle bottom). Importantly, the number of peptides for which high precision relative abundances of peptide pairs was increased by ~80% using DREAMS. In another case we observed 1,446 *D. radiodurans* proteins (46.4% of the predicted *D. radiodurans* proteome) as peptide pairs and also obtained good precision abundance ratios in a ***single*** DREAMS LC-FTICR analysis using our AMT-based approach.

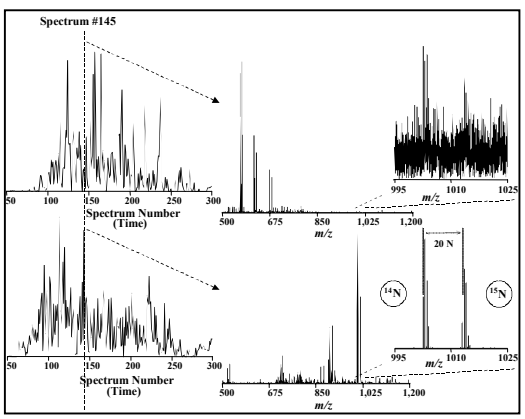

Fig. 4.3.2 Total ion chromatograms (TIC) of capillary LC-FTICR analyses, and typical mass spectra acquired during the "normal" spectrum acquisition and the alternating DREAMS spectrum acquisition for analysis of peptides from a mixture of natural isotopic abundance and ^{15}N-labeled mouse B16 cells. Top left: TIC reconstructed from the normal FTICR mass spectra. Bottom left: corresponding TIC obtained using DREAMS for which species having relative abundances >10% were ejected before ion accumulation. The mass spectra (center) show the ejection of the major species in the top spectrum from the one shown on the bottom. The detail (right) shows that a large gain in sensitivity and S/N is obtained for measurement of a specific peptide pair providing improved qualitative and quantitative results from the analysis.

Protein complex characterization at the intact-protein level. We have previously demonstrated the use of CIEF-FTICR to obtain broad proteome measurements for *E. coli* [Jensen, Pasa Tolic et al., 1999; Smith, Pasa-Tolic et al., 2001]. CIEF is an attractive separation technique for proteomics since it can handle extremely small sample sizes, has enhanced speed and resolution, can be easily automated, and provides high sensitivity when combined with MS arising from the natural concentration effect (a factor of 50-100 fold) associated with focusing. After focusing, protein bands are mobilized to the MS analyzer. PNNL has shown previously that analytes having pI differences as small as 0.004 pH units can be effectively separated [Shen, Xiang et al., 1999]. We have previously shown that >900 putative *E. coli* proteins could be detected in a single low resolution CIEF separation combined with FTICR [Jensen, Pasa-Tolic, et al., 2000; Jensen, Pasa-Tolic et al., 2000]. A limitation of these initial studies was that only a modest fraction of *intact proteins* can be initially identified based solely on mass measurements due to post-translational modifications that cause the mass of the protein predicted from its genome sequence to differ from its true experimentally measured mass. While MS/MS of smaller proteins is useful for identification, it is generally not well suited for the identification of large proteins, and is particularly

problematic due to present sensitivity constraints (an issue that we expect to be addressed by future research) [Shen, Xiang et al., 1999].

The situation is different when simpler mixtures from protein complexes are analyzed, particularly when information is combined with peptide level analyses. We have previously demonstrated the use of CIEF-FTICR for characterizing small protein complexes [Martinovic, Berger et al., 2000]. We have recently extended this approach to large complexes using capillary LC-FTICR [Lee, Berger et al., 2001] (see Appendix for preprint by Lee et al., PNAS in press). In this work we show that MW measurements for a finite set of proteins (already identified at the peptide level) is sufficient to infer the intact protein's mass. In the initial demonstration of this concept we analyzed the core proteins associated with the yeast large ribosomal complex, and found that the high mass measurement accuracies obtainable from FTICR combined with the identities of the 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). More importantly, analysis of the intact proteins revealed information on co- and post-translational modifications of the ribosomal proteins, including loss of the initiating methionine, acetylation, methylation, and proteolytic maturation [Lee, Berger et al., 2001]. Intact proteins not identified by this approach can potentially be identified by other methods (e.g. MS/MS) [Jensen, Pasa Tolic et al., 1999], and "data-directed" MS/MS analyses of such unidentified proteins is feasible [Li, Masselon et al., 2001].

4.3.1 Crosslinking of Protein Complexes

Chemical crosslinking is a widely-used technique in protein biochemistry [Wong, 1991; Mattson, 1993]. Crosslinking reagents contain a linear arm, of variable length and hydrophobicity, connecting two reactive groups, each of which is specific towards a particular site in a protein, such as primary amines or thiols. The elucidation of protein complex information obtained from crosslinking classically relies on throughput-limited techniques such as gel electrophoresis and N-terminal sequencing. Recently, however, mass spectrometry has been combined with crosslinking to provide a powerful set of tools for studying protein structure [Havel, 1979; Cohen, 1980; Young, 2000] and interactions [Rappsilber, 2000]. Determining the regions of two proteins in a complex that are crosslinked provides information on the identities of the proteins, as well as the geometry of their interaction in the complex [Lacroix, 1997; Rappsilber, 2000; Wallon, 2000]. That is, the fact that two peptides resulting from proteolytic digestion of a complex are crosslinked suggests that the regions of the proteins from which the respective peptides were obtained must have been neighbors in the complex.

Just as tandem MS can provide partial amino acid sequence for a peptide [Eng, 1994; Mann, 1994], it can provide similar information for crosslinked peptides resulting from crosslinking and digestion of a protein or protein complex. In a recent experiment at ORNL, bovine ribonuclease A was crosslinked using the reagent BS^3. The results of this experiment, and interpretation of the MS-MS and MS-MS-MS analysis of this crosslinked peptide pair, are described in the Preliminary Results section of the **Bioinformatics and Computation Subproject**. Briefly, cleavages of peptide bonds as well as an amide bond in the crosslinker were observed. A robust series of sequence ions was observed in the MS-MS-MS experiment, allowing identification of one of the crosslinked peptides, while the mass of the other (smaller) peptide could be inferred from the intact mass of the crosslinked pair, and from m/z offsets in the observed fragments from the larger peptide. This example illustrates several points relevant to MS analysis of crosslinked proteins. The MS-MS fragmentation of crosslinked peptides appears to be largely the same as in non-crosslinked peptides; the amide bonds between residues do fragment. However, the amide bonds between the crosslinker and the peptides also appear to fragment. Classical rules of interpreting tandem mass spectra of peptides are therefore still valid for crosslinked peptides, although they must be expanded. However, the interpretation is more difficult, and requires the development of computational tools described in the **Bioinformatics and Computation Subproject** if high throughput is to be achieved [Chen, 2001; Pearson, 2002]. To overcome experimental difficulties of detecting low-abundance crosslinked peptides in complex mixtures containing mostly non-crosslinked peptides, we propose a method using biotinylated crosslinking reagents to allow affinity enrichment of crosslinked peptides.

Synthesis of affinity-labeled crosslinking reagents. Figure 4.3.3 illustrates a reaction scheme employed at ORNL for preparing the biotinylated crosslinking reagents that will be vital to the success of this project. Two commercially available reagents, biocytin hydrazine and BS^3 (n=6; **Fig. 4.3.3**), were simply mixed in a 1:2 ratio in DMSO. Both the amine and hydrazine moieties react with the sulfo-NHS ester. The electrospray-Fourier transform mass spectrum of the reaction products showed three prominent signals in the mass spectrum, corresponding to different salt forms of desired product. A peak at m/z 1053.317 agrees with the predicted empirical formula ($C_{40}H_{61}N_8O_{19}S_3$) of the protonated form of the desired product within 0.003 Da, while peaks at 1075.298 and 1097.280 agree within 0.005 Da with the mono- and di-sodium forms of the desired product, respectively. The reaction mixture was also analyzed by reverse phase HPLC using a C_{18} column and eluting with a gradient of 0.05% trifluoroacetic acid and acetonitrile and detecting the products with a photodiode array detector (monitoring at 215 nm). By running the reaction with an excess of BS^3, the major impurity was the hydrolyzed product. The partially hydrolyzed product can be converted to the sulfo-NHS esters with DCC and sulfo-NHS. Additional strategies for synthesis and application of novel crosslinking reagents, including stable-isotope-labeled reagents, are being conducted by Dr. Malin Young and co-workers at SNL.

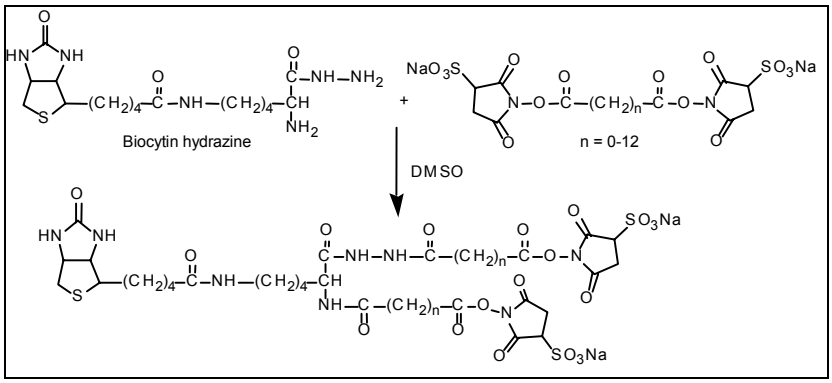

Fig. 4.3.3 Reaction scheme for producing biotinylated crosslinkers from commercially-available reagents.

4.4 Research Design and Methods

4.4.1 Increase the throughput, sensitivity and dynamic range of FTICR mass measurements (PNNL)

The high throughput characterization of protein complexes will benefit greatly from improved sensitivity and will be required to characterize low abundance proteins complexes. Achievable sensitivity tends to be a key factor defining data quality, and practical sample size requirements often dictate experimental practicality (e.g. in conjunction with high-throughput sample generation, to understanding how the composition of complexes changes with environmental conditions or cell state). Thus, a practical high throughput analysis capability will benefit significantly if the necessary protein quantity for each complex could be obtained from a 10 or 100 mL culture, as opposed to thousands of 10 L cultures. While the ability to work with such small samples also depends on the sample handling and processing conditions, it is obvious that improvements in overall analytical sensitivity are generally advantageous and essential to reduce the biomass that needs to be generated and processed to produce each sample for characterization.

Similar considerations apply for advances in dynamic range. An increase in dynamic range allows faster analyses of complexes since lower resolution LC separations can be used (because it is often unimportant that low-level peptides are separated from high abundance species). Greater dynamic range enables detection of less efficiently ionized peptides, and thus improves sequence coverage (and better identifying sites of modifications as well as providing increased overall confidence in identifications). Identification of low abundance species potentially provides insights into more weakly associated proteins. As more information is assembled covering large numbers of complexes and replicate analyses, it will be increasingly feasible to use informatic approaches to distinguish proteins associated at low levels from those arising due to contamination due to sample processing. Indeed, such low level associations may provide insights into cellular localization (i.e. they may not be artifacts, but convey useful information). A significantly extended dynamic range is also desired for measurements based on

crosslinking (where the requirements for both sensitivity and dynamic range are substantially greater than with other approaches), or in order to obtain information on lower level protein modifications relevant to biological function. Finally, increases in the dynamic range of measurements will correspondingly increase the quality of the data generated for lower abundance complexes, and will therefore increase the quantitative utility of the data. Such improvements will e.g., allow more definitive determination of whether detected proteins are constituents of a given complex or non-specifically associated species, since the ability to make such determinations will largely be based upon analysis of large sets of data obtained under multiple conditions.

The tasks under this Aim seek to advance the throughput, sensitivity, and dynamic range of protein complex measurements within the context of the AMT approach.

Task 1: Increase the sensitivity of FTICR mass measurements to enable characterization of lower abundance protein complexes and smaller sample sizes. As noted earlier, sensitivity is of great significance for high throughput protein complex characterization, and will become more important as sample processing procedures are refined, and advances such as the use of microfabricated devices are implemented (see **High Throughput Complex Processing Subproject**). PNNL has previously developed ESI-FTICR instrumentation that has achieved very high sensitivity based upon improvements to the ESI source design, high efficiency ion optics, improvements in FTICR signal amplification, and the introduction of the electrodynamic ion funnel [Belov, Gorshkov et al., 2000; Belov, Gorshkov et al., 2000]. The latter development alone provided an ~15-fold gain in ion current transmitted to the FTICR trap [Shaffer, Tang et al., 1997; Shaffer, Prior et al., 1998] allowing detection limits of ~30 zeptomoles (~18,000 molecules) for small proteins [Belov, Gorshkov et al., 2000]. Furthermore, the analysis showed that most of the remaining losses were associated with either ion formation from solution or the transfer to the first vacuum stage (i.e. through the ion inlet). This highlighted the potential for large gains under conditions where the delivery of analyte molecules to the ESI source is larger than can be efficiently ionized [Smith, Loo et al., 1990; Gale and Smith 1993; Wahl, Goodlett et al., 1993].

Our aim is to develop a new concept for increasing both the overall electrospray ion production rate and the efficiency of ion transmission and analysis. These gains are also anticipated to allow quantitative absolute abundance measurements to be obtained over a broader dynamic range (see Aim 3), and largely serve to eliminate ESI "suppression" effects that can cause significant bias in the analysis of mixtures.

It is now well established that ESI sensitivity increases as liquid flow rate to the source decreases [Goodlett, Wahl et al., 1993; Wahl, Goodlet et al., 1993; Wilm and Mann, 1994]. The greater efficiency at low flow rates results from the smaller droplet size and the similar levels of droplet surface charging by ESI. Overall ESI efficiency is expected to be ~100-fold greater for the use of 15 μm i.d. capillaries compared to the 150 μm i.d. capillaries as used for our initial capillary LC-FTICR studies of complex protein digests [Shen, Tolic et al., 2001; Shen, Zhao et al., 2001]. We have recently conducted analyses of complex peptide mixtures using capillary LC columns having i.d.'s as small as 15 μm, and shown that the sensitivity continues to increase inversely with eluent flow rates to as low as 20 nL/min. In that work we extensively examined the resulting signal intensities for mixture components of different peptide mass as a function of column diameter (or LC flow rate) and injected sample quantity. Typical signal responses for a few peptides as a function of the total sample mass injected using a 30 μm i.d. capillary is shown in **Fig. 4.4.1**. These data (and many similar results) show that above a certain sample size signal response can become non-linear. However, we expect that robust ESI based studies of protein complexes (e.g. using an automated system) will utilize ~50 μm i.d. capillaries (or in microfabricated devices of similar dimension) having flow rates of ~ 200 nL/min, and thus operate where the achievable electrospray ionization efficiency is ~10-fold lower than for similar separations using 15 μm i.d. capillaries. The primary reason for this is the lack of robust performance that makes operation problematic with lower "nanospray" flow rates.

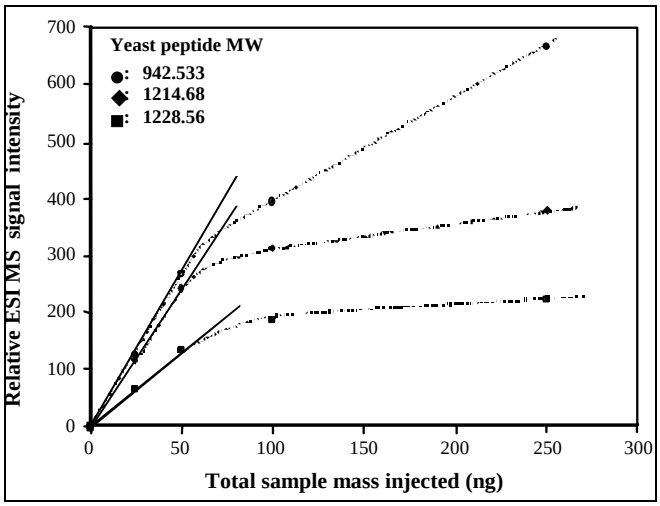

Fig. 4.4.1 ESI-MS peak intensities vs. sample size for three more abundant peptides obtained using a 30 μm i.d. packed capillary. For sample sizes below a given level MS peak intensities (or peak areas) are generally in the regime where they increase linearly with sample size, providing the best quantitation. The maximum sample size avoiding such suppression effects depends upon the mixture eluting from the LC at any given point, but is generally <75 ng for 30 μm capillaries, and <200 ng for 50 μm i.d. capillaries based upon results with microbial systems. This is a sample size regime generally useful with capillary LC-FTICR due to its high sensitivity.

We have recently shown that the overall ESI efficiency can be improved by distributing the LC effluent to multiple ESI emitter "tips", and the achievable total ion current increases in direct proportion to the number of emitters [Kim, Tang et al., 2001; Tang, Lin et al., 2001]. This approach benefits from the use of multiple capillary inlets from the atmospheric pressure ESI source in conjunction with a modified version of the electrodynamic ion funnel incorporating a novel "jet disrupter" [Kim, Tang et al., 2001]. However, the gains achieved initially were limited to only ~3-fold. The sensitivity gains were limited by: (1) space charge driven expansion of the ESI plume that lowered the current transmitted from the ESI emitter into the MS interface, and (2) the transmission from the interface into the rest of the analyzer, which is limited by the aperture connecting these regions.

In this task our approach will be to distribute the capillary LC effluent to a microfabricated ESI emitter array located in a reduced pressure region (~10 to 100 Torr), and where the losses due to the typically small (~400 to 500 micron diameter) inlet aperture from atmospheric pressure sources are avoided. The reduced pressure would decrease the pumping requirement, allowing a much larger diameter aperture between the ESI emitter array and the ion funnel (which can function at pressures up to at least 20 Torr [Tolmachev, Kim et al., 2001], and also contribute to more reliable/robust performance. It is worthwhile noting that previous attempts to implement low pressure ESI have failed due to less efficient droplet desolvation. In this effort we will use an extended ion funnel to extend the droplet desolvation period while also effectively confining ions to eliminate losses. Desolvation will be further aided by the small size of the individual droplets produced at the low flow rate to each emitter, and by the addition of heating to the region (if needed). (To provide sufficient heating for desolvation it may be necessary to concurrently cool the LC effluent flow and/or the ESI emitter array to avoid excessive heating of the liquid.) To avoid electrical breakdown at the reduced pressure, we will explore two promising approaches. In the first, the ESI will be conducted in a region of locally higher pressure due to the use of a sheath of high dielectric gas (SF_6) introduced around the emitter array. This gas can be heated if necessary to assist evaporation of the very small droplets initially produced from the microfabricated ESI emitter array. As a second alternative, we will explore the use of another approach to prevent electric breakdown. It has recently been suggested that electrosprays can be established from a microfabricated emitter, without the need of a separate counter electrode, by the use of an ESI emitter array electrically coated with a dielectric layer of conductive plastic [Shultz, Corso et al., 2001]. By applying the high voltage to the liquid and by grounding the conductive plastic, it was shown that an electric field could be established across the dielectric layer that was sufficient to generate stable electrosprays. The lower pressures we expect to use in the ESI source region will also allow the use of larger ion transmission apertures at the bottom of the ion funnel device, allowing larger ion currents to be transferred to the MS analyzer.

Thus, we expect to both substantially increase the raw analyte ionization efficiency and simultaneously increase the ion transmission efficiency. This development will provide substantial improvements in sensitivity and dynamic range when taken in conjunction with the DREAMS approach. Not only will it be possible to fill the FTICR ion trap at least 10-fold faster, but a similar gain will also be obtained for the DREAMS spectrum. The increased signal levels would also generally increase the accuracy of measurements obtained with Q-TOF type instrumentation (see Task 3), which are often signal (ion count) limited. A more subtle gain that we expect to realize is an increase in the concentration range over which ESI response is linear and for which highly

quantitative measurements can be obtained (see Aim 3). Success will draw on our experience with microfabrication [Lin, Xu et al., 1998; Lin, Matson et al., 1999; Wen, Lin et al., 2000; Tang, Lin et al., 2001], as well as those at ORNL.

Task 2: Increase the dynamic range of protein complex measurements. Increasing the dynamic range of MS measurements effectively allows increased throughput since the requirement for separating more efficiently ionized (or more abundant) peptides from less efficiently ionized (or less abundant) peptides is reduced. The DREAMS approach for obtaining extended dynamic range utilizes the information from the preceding mass spectrum to remove the most abundant species during the next filling of the FTICR ion trap. In principle, this allows the full dynamic range of the FTICR to be used for the subsequent spectrum. Initial results indicate that the overall dynamic range of measurements was increased by at least an order of magnitude.

In this task we will extend DREAMS to further increase dynamic range and to assist absolute abundance measurements of protein complexes (see 4.4.3). As initially demonstrated the DREAMS approach used different ion accumulation times for the normal and subsequent DREAMS spectra obtained continuously, in alternating fashion, during the separation. However, these ion accumulation times were fixed, at a short period for the normal spectrum (since the FTICR trap is more rapidly "filled" with ions) and longer period for the DREAMS spectrum (since the major species have been ejected and the ion current from the ESI source is now smaller). The best dynamic range in measurements is achieved when ions accumulate in the FTICR trap to near its capacity. However, in the course of LC separations the relative ion production rates for peptide mixtures can vary by over two orders of magnitude. Thus, a fixed ion accumulation time will result in one of three possible outcomes: (1) a less than optimal FTICR dynamic range, (2) FTICR trap "over-filling" where the accuracy of mass measurements is lost or reduced due to space charge effects, or (3) the loss of information on absolute ion production rates (i.e. when conditions cause ion removal in some fashion so as to avoid over-filling [Belov, Gorshkov et al., 2001]).

We will develop the capability to actively adjust ion accumulation times for *all* spectra. This step will allow better use of the full dynamic range in each spectrum, and further expand the overall dynamic range. To set the accumulation time we will use a fast "pre-spectrum" measurement to determine the ESI ion production rate. Preliminary experiments have indicated that the best approach is to use a short ion accumulation event in the external ion trap followed by ion transfer to the FTICR trap and their measurement. The use of "side-kick" ion trapping [Caravatti, 1988] for this allows immediate ion cyclotron excitation, detection of a low resolution "pre-spectrum", and automated selection of the accumulation time for the next spectrum in <300 ms. We will then use the information from the normal spectrum to both automatically select peaks for removal in the subsequent DREAMS spectrum *and* to set the longer ion accumulation time to be applied. These efforts will result in additional dynamic range gains when ion production rates are low during separations due to the accumulation times being longer, and will generally also result in correspondingly longer accumulation times for the DREAMS spectra. Thus, implementation of this approach will result in mass spectra during LC separations in which the accumulation times are *optimized for each normal and DREAMS spectrum*, considerably enhancing the dynamic range. Additionally, efforts to increase ionization efficiency under Task 2, and allow the FTICR trap to be filled with ions faster, will result in further improvements in both throughput (due to faster spectrum acquisition) and dynamic range (when limited by maximum accumulation times). We expect that by the end of year two these efforts will result in substantial increases in dynamic range, sensitivity and speed for characterizing protein complexes. These efforts will build upon expertise and resources in the EMSL at PNNL for the development of software and improvements in the computer control of FTICR instrumentation.

Task 3: Explore an alternative AMT based approach using potentially higher throughput Q-TOF technology. In this task we will explore a version of the AMT approach with the potential of significant gains in throughput relative to the FTICR approach using less expensive Q-TOF instrumentation. While FTICR technology provides the extremely accurate mass measurements that are the basis for proteomic measurements using AMTs, the reduced complexity associated with protein complexes may allow the approach to be implemented using instrumentation providing somewhat lower mass measurement accuracies (MMA), but with the advantages of lower cost and higher throughput. Using Q-TOF technology at PNNL we have achieved MMA of ~10 ppm. The MMA obtained using Q-TOF instrumentation are lower for low abundance peaks due to low ion

counts that do not allow accurate definition of peak centroids. This limitation can be mitigated by stronger signals (i.e. increased sensitivity). Indeed, we have significantly improved the routinely achievable MMA from the PNNL Q-TOF by adding an electrodynamic ion funnel to the interface to provide increased ion transfer efficiency from the ESI source.

Efforts under this task will determine if the MMA provided by Q-TOF technology, combined with LC elution time data, provides for the reliable identification of proteins from protein complexes. In this effort we will also explicitly use the redundant information obtained by the expected detection of multiple peptide fragments from each protein to determine the protein identity when MMA and LC elution time data is insufficient for protein identification. Thus, in contrast to the use of FTICR, which provides extremely high confidence for each peptide/protein identification, the use of Q-TOF instrumentation will more heavily depend upon the use of both LC normalized elution time (NET) data and a statistical determination of protein identifications exploiting the multiple peptides we expect to detect from each protein. In essence, this approach is an extension of the peptide matching approach routinely used for identification of single proteins, but augmented by the use of elution time information so that it allows the set of possible peptides resulting from mass measurements to be significantly reduced. For example, the three peptides QTFEAAILTQLHPR, TLHSLTQWNGLINK, and LLFLVGTASNPHEAR all have masses of 1605.86264 Da, but have much different LC retention times, and can thus be distinguished. The confidence for protein identification (and the "assignment" of detected peptide masses) will be based upon the expectation that multiple peptides will be observable for each protein component of the complexes. To do this we will modify our analysis software and data processing to implement the above approach. The efforts under this task potentially provide the basis for higher throughput than feasible using FTICR due to the much greater spectrum acquisition rates for Q-TOF instrumentation, and would be further advanced by developments that increase the ESI sensitivity, as well as on-going efforts elsewhere that promise to continue to improve the MMA obtainable using Q-TOF instrumentation.

4.4.2 Approaches for Improved Throughput for MS of Complexes (ORNL, Sandia)

Task 1: Combining crosslinking strategies with MS identification and characterization of protein complexes. This task focuses on integrating chemical crosslinking, affinity purification, MS, and computational techniques for identifying and characterizing protein-protein interactions. By forming a crosslink between two neighboring proteins in a complex, the interaction between these two proteins is "locked in" and can be maintained for analysis. The generality of the strategy derives from the potential ability to crosslink, digest, and analyze all protein complexes in a complex mixture, such as a cell lysate, without the need to engineer a "bait" protein or an antibody. The versatility of this strategy lies in the wide variety of crosslinking reagents reported in the literature (or that can be synthesized), and in applying a wide range of reaction conditions via robotics and combinatorial methods.

Using an affinity tag on a protein labeling reagent to reduce the complexity of a peptide mixture is one basis of the isotope-coded affinity tag (ICAT) strategy for quantitative proteomics [Gygi, 1999; Griffin, 2001]. A tag attached to a crosslinking reagent has been described that will provide a unique mass spectrometric signature in tandem mass spectra of crosslinked peptides [Back, 2001]. The key element in our proposed strategy combines these ideas in the form of an *affinity-tagged crosslinking reagent*, which will allow isolation of crosslinked peptides from protein complexes, preserving the crosslinking information while greatly reducing the complexity of the mixture.

The first step will be to crosslink to one another the adjoining proteins involved in complexes. The crosslinking reagent will have three functional groups, as shown in **Fig. 4.4.2**. Two of these functional groups will be reactive or photoactivatable and will thus crosslink the proteins. The third functional group will be a biotin affinity tag.

After crosslinking, the sample will be dialyzed to eliminate unreacted crosslinking reagent and endogenous biotin. Then, a proteolytic digest will be performed. The products of this digest will be two classes of peptides: (1) a large number of non-crosslinked peptides, and (2) a much smaller number of crosslinked peptide pairs. The peptides in any given intermolecularly crosslinked pair will be fragments from a pair of proteins that were originally neighbors in a complex. These crosslinked peptide pairs incorporate the biotin tag, which will allow them to be affinity captured on an avidin column while the peptides from class (1) are washed away, greatly simplifying the mixture. The rosslinked peptide pairs of class (2) can then be eluted from the avidin column, and analyzed using ThermoFinnigan LC-MS-MS instrumentation.

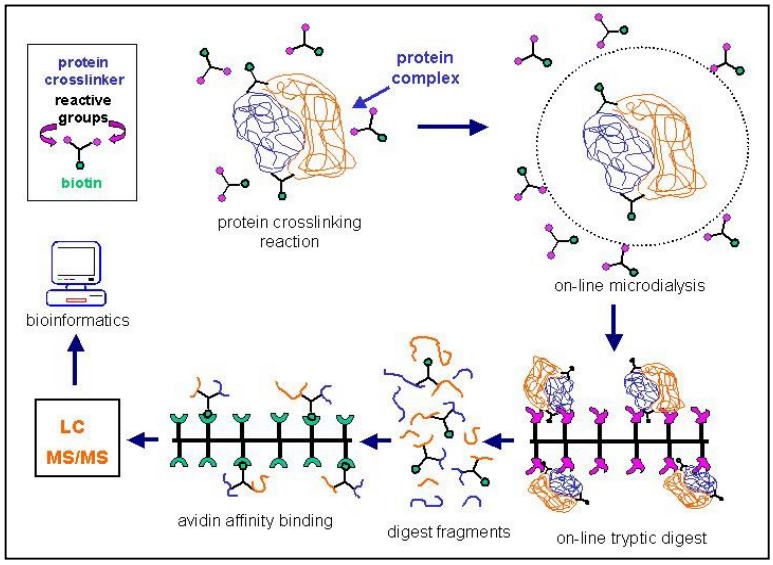

Fig. 4.4.2 Schematic of crosslinking strategy using trifunctional crosslinker to stabilize complexes and permit rapid isolation of crosslinked fragments from complex mixture.

One "trifunctional" crosslinker that incorporates two reactive groups and a biotin tag is commercially available (Sulfo-SBED from Pierce) but we propose to synthesize more suitable reagents. At the 2001 ASMS meeting in Chicago, Michelle Trester-Zedlitz of Brian Chait's group (Rockefeller University) presented a solid-phase peptide-style synthesis of a crosslinker similar to those we are describing. We propose a more flexible solution-phase synthesis strategy.

Trifunctional crosslinking reagents with spacers arms ranging from 6 Å to over 30 Å will be synthesized that incorporate two protein reactive groups and a biotin affinity tag. Since lysine residues (which contain a terminal primary amine) are usually the most abundant amino acid in proteins, the first crosslinking agents will contain functional groups that specifically react with amines, such as succinimidyl derivatives of carboxylic acids and imidoesters. Initially, sulfonated *N*-hydroxysuccinimidyl (sulfo-NHS) esters will be prepared from dicarboxylic acids and *N*-hydroxysulfosuccinimide by coupling with *N,N*-dicyclohexylcarbodiimide (DCC) in dimethylsulfoxide (DMSO) or dimethylformamide (DMF) [Muller, 2001]. For crosslinking reagents with short linker arms (6-8 Å), aspartic acid or glutamic acid will be coupled to biotin (using DCC in DMF) or the *p*-nitrophenyl ester of biotin [Bodanszky, 1977]. The resulting amide linkage is hydrolytically stable, as opposed to the imidoamides, formed from the reaction of an amine with an imidoester, which hydrolyze at high pH. The dicarboxylic acids will be converted to the sulfonated *N*-hydroxysuccinimidyl esters by standard methods. For crosslinking reagents with longer linker arms (9 to over 30 Å), biocytin hydrazine (available from Sigma), which contains two reactive amines separated by ca. 4.3 Å, will be derivatized with sulfonated *N*-hydroxysuccinimidyl esters of dicarboxylic acids. Since dicarboxylic acids with zero methylenes, i.e., oxalic acid, to 12 methylenes, i.e., 1,12-dodecanedicarboxylic acid, are commercially available, spacers arms ranging from 9 to over 30 Å can be prepared by this method. Depending on the results obtained from these experiments, additional crosslinking reagents will be synthesized. For example, a thiol-reactive bis *N*-maleimide derivative could be prepared by reacting a sulfo-NHS ester of maleimidoalkylcarboxylic acid (commercially available) with biocytin hydrazine.

It is generally recognized that conditions for crosslinking reactions tend to be specific to a given protein or complex. Therefore, no single set of conditions will be ideal for crosslinking all of the protein complexes in a mixture such as a cell lysate. A successful experimental strategy will require subjecting a given lysate to a variety of crosslinking conditions, including pH, crosslinker and protein concentrations, temperature, reaction time, etc. Existing robotic equipment at ORNL allows such a variety of conditions to be implemented. An additional

complication involves interferences from primary amines in non-protein components that can react with lysine-reactive crosslinking reagents. It may be possible to overcome this problem by increasing the crosslinker concentrations. Endogeneous biotin and biotinylated species will compete with the biotinylated crosslinked peptides for sites on an avidin affinity column, while endogenous streptavidin can bind biotin-labeled crosslinked peptides. If these problems arise, we will remove these species prior to crosslinking using affinity techniques.

The "workhorse" tool for the MS analysis is expected to be HPLC ESI MS/MS implemented on Finnigan LCQ instruments at ORNL; however, other mass spectrometric techniques are available for cases arising during initial phases of this project for which HPLC ESI MS/MS is not the optimum tool. MALDI-TOF-MS (Hurst lab) and ESI-FTICR (Hettich lab) can be brought to bear for these measurements.

Sensitivity and dynamic range issues will determine cell culture volumes that must be employed to obtain successful identification of protein complexes. For example, an *E. coli* cell with a volume of ~$3x10^{-12}$ cm^3 will contain about 15% protein by weight [Lehninger, 1975], or ~$5x10^{-13}$ g protein. If 1% of all the different types of complexes in a sample are crosslinked under a given set of reaction conditions, and further assuming 10% crosslinking efficiency for a protein complex for which conditions are optimum, then about 0.1% of the protein complement of a cell, or ~$5x10^{-16}$ g, will be involved in crosslinks. If each protein subunit yields 20 peptides (assumed to be of equal mass) on digestion, and one of these peptides is involved in a crosslink, then about 0.005% (~$2x10^{-17}$ g) of the protein complement will be present as crosslinked peptide pairs after digestion. For a capillary HPLC column on our LCQ system, a 1 picomole quantity of a peptide gives abundant MS/MS signal. For a crosslinked peptide mass of ca. 3000 Da, 1 pmol = $3x10^{-9}$ g. To yield this amount of peptide would require about $2x10^8$ microbial cells, multiplied by the number of different complexes contained in the 1% of complexes that are crosslinked. That is, if 10 different complexes were crosslinked, $2x10^9$ cells would be required as a starting material, translating to a reasonable 100 mL of a typical *E. coli* cell culture (10^7 cells/mL). Protein complexes at lower concentrations will require correspondingly greater amounts of starting material to be reliably detected by our technique.

Experiments will progress from initial demonstrations of crosslinking, digestion, affinity separation, and MS identification of components of a model protein complex that is relatively easily isolated from *S. oneidensis* or *R. palustris*. A progression toward more realistic systems will involve adding increasing amounts of other proteins and interfering compounds to the reaction to simulate matrix effects that one would encounter in an actual cell lysate or fraction. These experiments will provide information on the level of spurious crosslinking between 'unrelated' proteins. Once mixture effects are understood, two parallel paths are planned as this research component is evaluated for transition into the **Center for Molecular and Cellular Systems Subproject**. The first path will be experiments to characterize discrete isolated protein complexes isolated from *S. oneidensis* and *R. palustris* in the **Center**. The second path will be exploration and validation of the affinity-labeled crosslinking technique for surveying a variety of complexes in a mixture. This path is the “protein complex” analog to “shotgun” proteomics approaches [Washburn, 2001] that do not require separation of the proteins in a sample prior to digestion and MS analysis.

We will work toward the eventual goal of an integrated system for acquiring the experimental data. The dialysis, digestion, and avidin binding steps could all be performed in sequential modules (columns) of a flowing system, and therefore seamlessly integrated with LC-MS-MS instrumentation. Integrating the various steps in the procedure would greatly increase throughput, capabilities for automation, and scalability, directly addressing GTL requirements. Interpreting mass spectra resulting from digestion of affinity-labeled crosslinked protein complexes will require several new computational tools, described in the **Bioinformatics and Computing Subproject**. This work will be performed in collaboration with Sandia National Laboratories.

Task 2: Protein chip/ion trap array read out system for identification of protein-protein complexes. We propose to develop means to read out protein capture arrays by electrospray tandem mass spectrometry (ESI-MS/MS).

Protein arrays have recently been shown to be a viable tool to identify protein-protein interactions [Zhu, 2001].

This approach to screening for specific protein-protein interactions benefits from the same advantages as widely-used DNA microarrays for mRNA expression analysis, namely high-throughput parallel, quantitative microscale analysis. However, both approaches require the use of extrinsic labeling, but protein microarrays cannot easily exploit the simple labeling and detection (read out) schemes used for DNA arrays. The variety of proteins that may interact with each other, the range of binding constants, and the possibility of post-translational modifications, each presents a significant challenge for array read out. A means to read out these interactions directly using mass spectrometry, which requires no extrinsic labeling or post capture on-chip chemistry, would provide an important advance in the use of protein arrays for protein-protein interaction analysis. This technology is obviously synergistic with research aimed at producing "bait" proteins and scFv antibody reagents (see **High Throughput Complex Processing** and **Center for Molecular and Cellular Systems Subprojects**).

Currently, two ionization sources can generate gas-phase ions from proteins for analysis by MS, viz., MALDI and ESI. Of these two, the apparent best choice for protein array read out would be MALDI-MS, which is a surface analysis technique. In ESI-MS, analytes are introduced to the mass spectrometer in a liquid solution, which presents the question of how to transfer the captured material from the array into solution. MALDI-MS approaches to protein chip read out are, in fact, currently being exploited by two different companies, viz., Ciphergen Biosystems, Inc. (Fremont, CA) and Intrinsic Bioprobes, Inc. (Tucson, AZ). The commercial availability of MALDI-MS protein chip products speaks of their utility. Nonetheless, the use of MALDI-MS for chip read out presents analytical limitations. There is a low number density of analyte at any small area on a particular array spot illuminated by the laser beam to generate ions (which negatively impacts detection levels). Detection levels in MALDI-MS precipitously decline above a molecular mass of about 15 kDa severely limiting the range of proteins that can be analyzed directly. Time consuming enzymatic digestion methods are needed for generating low mass peptides that are more amenable to detection when larger proteins are analyzed. These digestions are also needed to generate peptides for protein identification by peptide mapping. Mass accuracies in MALDI-MS are usually no better than about 0.01% ($\pm$ 6 Da for bovine albumin, ca. 66000 Da). Finally, analysis of the arrays requires removal from the native liquid environment within which the interactions occur, the application of a chemical matrix to facilitate desorption and ionization, followed by a drying step.

The ability to apply ESI-MS rather than MALDI-MS for chip read out would provide the same type of data, while overcoming limitations of MALDI-MS. ESI-MS introduces the sample to the mass spectrometer as a liquid solution, allowing one to sample the components interacting at each point on an array under a liquid environment. All the captured material on any interaction spot can potentially be collected and directed to the mass spectrometer, not just the very small fraction that interacts with the laser beam in MALDI-MS. ESI-MS does not have the same drop off in detection level as does MALDI-MS as molecular mass increases. Furthermore, up to a mass of about 60 kDa, even modest mass analyzers with nominal mass resolution can obtain mass accuracies as good or better than $\pm$0.002 % ($\pm$1.3 Da for bovine albumin). Even better mass determinations can be provided by higher performance mass analyzers such as the ESI-equipped 9.4 Tesla FTMS instrument at ORNL. Thus, proteins can be identified on the basis of high accuracy molecular mass determinations, and these measurements can be performed even when mixtures of proteins are present. Tandem mass spectrometry of the whole proteins eluted from the arrays can be used to generate data for protein identification by database searching [Mortz, 1996]. This fast (ms time scale experiment) "top-down" approach to protein identification eliminates the need for costly and time-consuming enzymatic digestion methods. Computational techniques described in the **Bioinformatics and Computing Subproject** for analysis of "top-down" proteomics data will be applicable to this research.

We have demonstrated the use of surface stepping sampling probe (SSSP)/ESI-MS for reading out arrays (unpublished data). This ESI-MS surface sampling system combined with proper MS instrumentation provides for a relatively simple, automated, and molecular mass and structure specific means of array read out. It is clear, however, that better sensitivity and truly high throughout require a change in thinking about the arrays for read out. The array designs used for optical or MALDI-MS read out are not ideal. The array should be designed specifically for ESI read out. We propose to develop an ESI array read out device making use of soon to be commercially available (from Advion Biosciences, Ithaca, NY) microfabricated arrays of ESI nozzles in which each nozzle is connected to a distinct well on the opposing side of the array. Using chemistries already established in our [Subramanian, 1999] and other laboratories, we will immobilize capture proteins in these wells (**Fig. 4.4.3**).

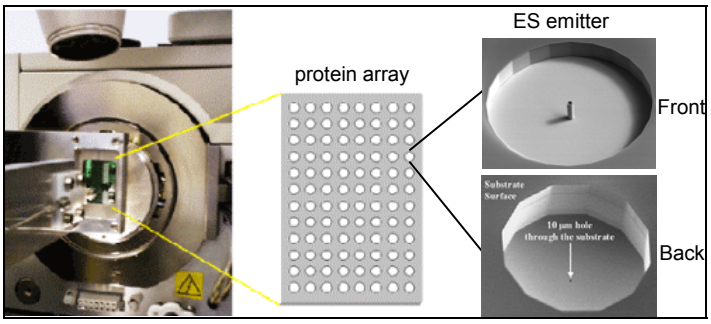

Fig. 4.4.3 Protein Chip Array being read on ES-MS/MS (left). Immobilized captured protein in a well (right).

In this arrangement, the capture array and electrospray emitter system will be one unit. Sampling, cleanup, elution, and spraying are integrated on one device. This should enable creation of densely packed arrays for ESI-MS read out, provide for very low detection levels (high concentrations of the eluted species, because of the minute, pL, void volume), and provide for fast (5-10 s/spot) and reproducible serial read out of the chip array. Onto the chips will be immobilized particular protein "baits" or scFv antibodies (from the **Center**) for which complex information is desired. Following incubation of the array spots with the samples of interest, tandem mass spectrometry of the captured whole proteins eluted from the arrays will be used to generate sequence tags for protein identification by database searching. We will work to develop rules (see **Bioinformatics and Computing Subproject**) of whole protein fragmentation and algorithms that permit sequence tags to be gleaned from tandem MS data acquired first on the higher performance FTMS instrument, then subsequently on lower cost ion trap instrumentation.

A second generation advance in this technology will be to mate microfabricated ESI capture arrays with microfabricated arrays of ion trap mass spectrometers for fast, parallel read out of the entire array (5-10 s/array). These read out times for an entire array will be as fast as those employing optical methods for array read out.

We have shown that ion traps with fundamental dimensions <1 mm can be used for analytical MS with no sacrifice in mass resolution over conventional-sized instruments [Kornienko, 1999]. We have also shown that a two-dimensional array of these miniature traps can be operated in parallel [Reilly, 2000]. The purpose of the proposed work is to develop a microtrap array for parallel ESI-MS analysis of protein microwell arrays. Such instrumentation could result in nearly a hundred-fold reduction in analysis time compared to serial analysis.

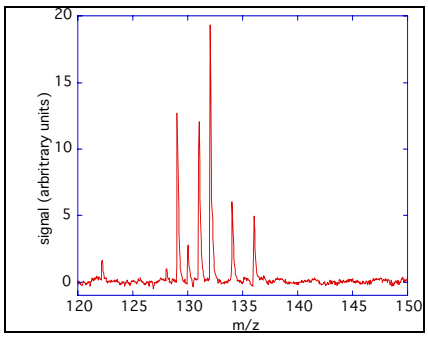

Fig. 4.4.4 Mass spectrum of xenon with 7-trap array

In an initial experiment, an array of seven traps was operated in parallel with a common detector. The ion traps were operated in parallel with essentially identical internal fields. A mass spectrum of xenon gas obtained with the 7-hole ion trap array is shown in **Fig. 4.4.4.** The major isotopic ion peaks are clearly resolved in this spectrum. These measurements were made with a higher detector time constant than for the single trap experiments so it is not possible to directly compare the mass resolution of the two configurations. However, the leading edge of the peaks in **Fig. 4.4.4** is comparable in slope to the earlier measurements suggesting that the resolution for the array is not appreciably degraded. The signal in **Fig. 4.4.4** was approximately twice that obtained with the single hole trap. We would not expect to achieve the full 7X in signal because the whisker array did not completely overlap the end cap apertures. The experiment demonstrated that an array of traps can be used to increase the ion storage capacity and signal to-noise ratio of microscale ion trap mass spectrometers.

To extend this concept to parallel electrospray analysis would require an atmospheric pressure interface to admit the electrosprayed sample from each well into the appropriate ion trap with suitable ion guides for focussing and solvent removal. A detector array would be employed to measure the ions as they are ejected from each trap as during the m/z scan. The initial accomplishment for this task will be to construct an electrospray inlet for the existing single 1-mm diameter ion trap MS and establish optimum conditions for protein mass spectrometry. The performance of the spectrometer would be evaluated in combination with prototype protein arrays. Subsequent efforts would be to replace the single trap with a 2-D array of ion traps, add a multiple spray input capability, and replace the single detector with a detector array so that each input could be analyzed individually.

High throughput is a high priority for Goal 1 of GTL. The results of this task offer a powerful combination of techniques for performing protein interaction measurements in parallel.

4.4.3 Develop Approaches for Obtaining Improved Protein Complex Identifications, more Accurate Absolute Protein Abundance Measurements and Improved Assessments of Data Quality (PNNL)

Task 1: Increase the FTICR mass measurement accuracy to enable faster and more confident protein identification using AMTs. Improvements in MMA allow identification of more peptides with greater confidence, and will result in improved confidence in identifications and/or higher throughput. Our calculations have shown that the ability to identify peptides continues to increase with MMA to at least 0.1 ppm [Conrads, Anderson et al., 2000]. It is well established that FTICR excels in this regard and that measurements are improved by the use of high magnetic fields such as used at PNNL [Marshall and Guan, 1996; Marshall, 2000].

In this task we will develop and implement a more robust and general approach for including mass standards in each spectrum to further enhance the MMA of FTICR measurements. To achieve a more robust internal calibration, we will develop a variation of the approach described by Muddiman and coworkers [Hannis and Muddiman, 2000; Flora, Hannis et al., 2001] who used a second ESI source which could be rapidly swapped with the primary source using a solenoid drive to introduce mass calibrants. An attraction of this approach, in contrast to approaches using completely independent ion inlets, is that the analyte and calibrant ions will have spatially identical locations in the FTICR cell, and thus be ideally suited for calibration (mass measurements are affected by any differences in magnetic or electric fields). A difference in our approach will be the use of a second source and inlet that is spatially distant from the primary ESI source, and which also has a smaller i.d. capillary inlet (so as to minimize the additional gas load to the MS interface region), and which will also be compatible with our anticipated ESI source developments described above (4.4.1). The second source will use a calibration mixture that will add 3 to 6 distinctive *known* peaks to each spectrum. The inlet will direct ions to the periphery of the ion funnel and will be offset from the primary ESI source so that their performance will be mutually independent, but after focusing in the ion funnel both will have exactly similar ion optic characteristics and thus provide an ideal situation for calibration. An important aspect of our design will be the use of electrical gating on the low vacuum side of the interface so that the ions from the calibration source will either be directed into the ion funnel or deflected from entry to the ion funnel. Initial results based on a modified version of PNNL's "jet disrupter" ion funnel design [Kim, Tang et al. 2001], have indicated that a dual inlet variation of this design should work extremely well for this application. Thus, we will be able to have rapid electrical control of the calibration ion inlet. In conjunction with the DREAMS technology, we will then develop the capability to introduce approximately *the same number of calibrant ions to each spectrum* (regardless of ion accumulation time for either the normal or the DREAMS spectra). The calibrant ion inlet time will be selected to provide low intensity calibration peaks in each spectrum so as not to degrade overall dynamic range. The operation of the ESI calibration source will be integrated with the overall DREAMS development, e.g. so as to avoid ejection of all calibrant species in the second DREAMS spectrum.

In this task we will potentially pursue developments expected to increase the achievable MMA. For example, we will exploit our recent understanding that the commonly applied (and most significant) correction to cyclotron frequencies needs to be altered since the shifts are not the same for all peaks, but in fact are more correctly proportional to the total ion population *minus* the ion population for the specific peak being perturbed [Masselon, Tolmachev et al., 2002]. We have also found that other corrections to the fast FT should likely be applied to closely spaced peaks [Tolmachev, Masselon et al., 2002]. Such developments should provide improved MMA, and allow us to exploit the additional improvements that are expected as more peaks in each spectrum can be confidently assigned as AMTs; each additional peak assigned with very high confidence can serve as an effective internal calibrant and improve the MMA for other peaks in the same spectrum. Thus, we envision that substantial overall improvements in MMA can be achieved by an iterative process. Our initial analysis has indicated that a MMA of <0.25 ppm is achievable [Masselon, Tolmachev et al., 2002], which would enable a significant extension in the use of AMTs and/or increased confidence in identifications. These improvements will be

manifested as either qualitative improvements in, or as increases in the levels of confidence associated with protein identifications.

Task 2: Develop methods for improved quantitation and providing measures of data quality. It is well recognized that MS peak intensities do not accurately reflect peptide (or protein) absolute abundances in a solution [Cech and Enke, 2000; Cech and Enke, 2001]. For example, analysis of an essentially "complete" tryptic digest for a protein will yield a spectrum where peaks for different peptides span more than an order of magnitude in relative intensity. The reasons for this are multifold, and include variations in electrospray ionization efficiency related to peptide composition (e.g. hydrophobicity) [Cech and Enke, 2001], and the variation in solvent composition during the LC gradient [Cech, Krone et al., 2000]. Similarly, selective losses in sample handling can depend upon peptide properties. Additionally, the effectiveness of enzymatic cleavage at some sites can vary due to peptide sequence (e.g. trypsin is often less efficient at cleaving between lysine and acidic residues) [Wilkinson, 1986].

In order to compare the relative abundances of protein complexes for two different states of the microorganism, stable-isotope labeling can be effectively applied and result in precisions that can be <10%. One general approach that we have found particularly effective is to conduct metabolic ^{15}N -labeling (i.e. culture the organism in ^{15}N enriched media [Conrads, Alving et al., 2001]). This approach allows ^{15}N and normal ^{14}N-labeled cell populations to be mixed prior to cell lysis and protein complex recovery, and will provide the best precision for measurements of the relative abundances of complexes between the two culture conditions. (Experience to date indicates that any ^{15}N-isotope effects on biological processes are minimal; if this proves not to be the case such effects can be mitigated by either reversing the application of the $^{15}N/^{14}N$ labeling and correcting for such effects, or by the use of a common ^{15}N-labeled calibrant added to two different experiments.) While the use of stable-isotope labeling in such comparative studies allows extremely precise measurements of changes in relative abundances between two samples, it does not provide good information on absolute abundances. Only semi-quantitative information is obtained from the peak intensity measurements, and significant improvements could be realized by effectively correcting such measurements for variations in ionization efficiency resulting from peptide composition, etc.

We will exploit the capabilities to significantly improve measurements of absolute ion abundances (or peak intensities), and which will also benefit from the increased dynamic range of measurements we expect to achieve (see 4.4.1). To help establish an absolute abundance scale, we will include a set of standard peptides (and a small protein) in each capillary LC-FTICR (or Q-TOF) analysis. These added components will provide "anchors" for the absolute quantitation, provide information on the effectiveness of the proteolytic digestion step, and enable the interpretation of measurements in terms of peptide (or protein) abundances/cell or sample mass. Our goal here is to improve the accuracy of the protein abundances derived from reversed phase capillary LC-FTICR analysis of peptides. Due to the very high sensitivity achieved and the inherent sample dilution occurring during an LC separation, very low concentrations of analytes are actually eluted to the electrospray ionization at any point. The linearity of ESI response for sets of peptides under such conditions has been demonstrated [Cech and Enke, 2001], and also observed in our capillary LC-MS proteome analyses using complex peptide mixtures when the maximum sample size is constrained (see Aim 1, Task 2). Thus, it appears that ESI ion suppression or saturation effects can be avoided by working in this regime (and where sensitivity is also maximized).

To increase the quantitative value of measurements (e.g. to better determine protein stoichiometries**)** we will develop an approach to correct for variations in ionization efficiency based upon information derived from the tryptic digestions of mixtures of equi-abundant (e.g. standard) proteins followed by their analysis under our typical reversed phase capillary LC conditions to obtain the peak abundance information for all the resulting tryptic peptides. The overall variations in ionization efficiency under relevant conditions arise substantially, and probably predominantly, due to peptide composition and the LC solvent gradient. We will therefore develop artificial neural network models in a fashion similar to how we have initially developed an LC elution time prediction capability (see Task 1). An attraction of this approach is that it will be possible to build an initial development data set based upon the use of as few as 20 to 40 proteins and the thousands of peptides resulting from their digestion with trypsin. Thus, we expect to develop a capability to correct peptide peak intensity data from our analyses using only a peptide's sequence and elution time. The abundances for each protein can be put

on an absolute abundance scale using the "anchor" information derived from the set of reference peptides to be included in each capillary LC-FTICR analysis. Thus, substantially improved measures of absolute protein abundances should be achievable.

The above approach will allow us to develop statistical measures of quality for each protein complex analysis based on abundances for the set of calibrant peptides. The set of calibrant peptides will include a set of stable-isotope labeled peptide pairs to provide an assessment of the quality of relative abundance measurements. The lower abundance peptides will have greater numbers so as to provide a better basis for estimating uncertainties. Multivariate regression methods will be used to fit a model function that includes mass spectral intensities (integrated over the elution profile) so that absolute abundances can be computed [Boggs, Donaldson et al., 1989]. Our aim is a statistically sound measure of data quality for each protein complex measurement.

4.4.4 Develop Approaches for Obtaining Improved Protein Identifications (ORNL, U. Utah, U. North Carolina)

Task 1: Linked LC-MS and LC-FTICR-MS studies for high-throughput peptide characterization. An LC-MS approach to be developed at ORNL involves the coupling of a single two-dimensional LC system with both quadrupole ion trap (QIT) and FTICR-MS instrumentation. The goal of this approach is to analyze the digest of a protein complex for accurate peptide molecular masses (FTICR-MS) and rapid tandem mass spectrometry (QIT) information in *a single LC-LC experiment*. The LC-LC system will employ (for example) a strong cation exchange column (500 um x 10 cm) followed by a separate reverse-phase C18 column (150 um x 15 cm). The LC-LC experiment would be conducted in the standard "MudPIT" mode, except that the eluent from the final reverse phase column would be split between the QIT (to conduct MS and MS/MS experiments on the peptides) and the FTICR-MS (to measure peptide masses under high resolution conditions). While some of the experimental conditions remain to be worked out (such as optimum column sizes and gradient conditions), this approach would permit the precise correlation of accurate peptide masses (measured to less than 1 ppm) with their corresponding tandem mass spectrometry data. Note that normal between-run and column-to-column fluctuations in the chromatograhic retention times become irrelevant, as the tandem mass spectrometry and high resolution mass measurement experiments are always conducted on the identical LC eluent. This would not only enhance the reproducibility of day-to-day analyses, but would also enable experiments to be conducted on organisms for which accurate mass tag libraries had not been generated, which would be particularly important for the rapid examination of "previously uncharacterized" peptides. Furthermore, this approach should be an efficient method for protein quantitation with stable isotope labeling techniques. We propose to evaluate the $^{18}OH_2$ labeling procedure, developed by Fenselau and co-workers [Yao, 2001], in which one sample is proteolytically digested in normal water, while the other sample is digested in $^{18}OH_2$. The high resolution mass measurement of the FTICR-MS would enable unambiguous peptide identification (even in the presence of the small isotopic shifts) whereas the QIT would rapidly acquire tandem mass spectrometry on all of the eluting labeled and unlabeled peptides.

Task 2: Development of new chromatographic-MS techniques for identifying protein-protein interactions. A proteomic approach has been demonstrated at ORNL for fractionating intact proteins by anion exchange fast protein liquid chromatography (FPLC), followed by off-line MS identification utilizing both "top-down" and "bottom-up" MS methods [VerBerkmoes et al., 2002; reprint in Appendix]. In the course of these studies, we have noted numerous instances where identical proteins elute under widely separated, non-adjacent preparative anion exchange fractions. Since the anion exchange FPLC is conducted under non-denaturing conditions (pH 8, 20 mM tris, 0 to 1 M NaCl), it is likely that some of the protein complexes that survive the cell lysis and protein cleanup steps are resolved in this fractionation. This might explain why some proteins elute in one particular fraction (as single species), and then the same proteins (as part of a multi-component complex) elute in a later fraction. If this process is reproducible over a range of experimental elution conditions, then it might be possible to exploit this approach to aid in elucidating protein partnering.

We have noted this elution of identical proteins in non-adjacent fractions in both microbial [VerBerkmoes et al., 2002] and yeast cell lysates. In fractionation of yeast proteins by anion exchange FPLC, two phosphoglycerate proteins elute in fractions 19 and 43, but not in the intervening fractions. By comparing the proteolytic peptides of

these two fractions with the high resolution ES-FTICR-MS mass measurements of the intact proteins, it was possible to verify the presence of phosphoglycerate kinase and phosphoglycerate mutase in both fractions. Both of these proteins were modified; the predicted N-terminal methionine of the mutase was absent, and the kinase revealed N-terminal methionine loss accompanied by addition of an acetyl group. The measured masses of the intact proteins were within 3 ppm of their calculated masses (with PTM's). This mass accuracy allowed verification that the larger protein (measured at 44,647.714 Da) corresonded to the modified phosphoglycerate kinase, and not yeast flavohemoglobin (calculated Mr = 44,646.1123 Da). It is interesting to note that these two phosphoglycerate proteins are closely related, and are employed in adjacent steps in the glycolysis pathway. Inspection of the recently-published yeast protein interactions [Gavin et al., 2002; Ho et al., 2002] does not reveal any detected interactions between the kinase and the mutase, although neither of these proteins were entry points for the interaction measurements. Our preliminary data imply that these two proteins are linked by interaction either with each other, or as part of a larger system. This example illustrates the importance of using multiple experimental approaches to elucidate protein interactions, and the need for bioinformatics tools to integrate data from these approaches.

We propose to evaluate more rigorously this method to identify interacting proteins. Work would begin with yeast studies to search for known protein partners, and then extend to known complexes from *S. oneidensis* and *R. palustris*. A variety of FPLC separation modes (anion exchange, cation exchange, pH) would be systematically studied to unravel as many protein-protein relationships as possible. We do not expect that this technique would replace the pull-down methods outlined previously, but rather would complement them by providing additional information on the subsets of protein-protein interactions within a given interaction map.

Task 3: Global approaches for protein identification by LC-MS. To address the more general issue of MS approaches for analysis of intact proteins, two different approaches have been developed: Global analysis of proteins fractionated by front-end column chromatography and one-gene-at-a-time analysis by tagging and affinity purification. Both approaches require new computational approaches for identification of open reading frames that contribute to encoding the protein of interest and to sort through the many possible contributions to protein mass other than just the predicted amino acid content [Baronov et al, 2001; Shah et al, submitted 2001] (See the Bioinformatics and Computing Subproject for a description of the computational activities being conducted at U. North Carolina in concert with these MS activities). There are two ongoing biological applications of these technologies in the Utah lab. One, initiated with a grant from the DOE genome program, is focused on yeast *Saccharomyces cerevisiae*. The yeast proteins are being fractionated by column chromatography and pooled fractions are being analyzed by LC/ESI/MS in order to get accurate masses of intact proteins. The same fractions are digested with trypsin and these masses are used to identify the genomic origin of the sequence [Moore, et al, 2002]. Differences between the mass of the intact protein and the mass predicted from its open reading frame then provide the starting point for understanding the real details of the pathway to the mature protein. The initial effort is on yeast mitochondria, which is a convenient source of about 10% of the total protein complement. The second effort is focused on characterizing the protein products from open reading frames in microbial genomic sequences that suggest unusual translation events such as two open reading frames linked by a stop codon or two overlapping open reading frames.

The Utah group will apply this new technology to the goals of GTL by aiding and augmenting the approaches of ORNL and PNNL. The primary expertise of the Utah group is molecular genetics with very extensive experience in genetic constructs for analyzing the expression of proteins from genes. A number of different, alternative affinity tags are engineered into genes of interest so that the protein products can be purified by simple affinity chromatography and can then be analyzed by ESI/MS in order to determine their mode of expression. This genetic tagging approach can be exploited to characterize molecular complexes from cells. Individual bait genes will be tagged so that their protein products will allow purification of the protein, and complexes containing them, by affinity to a matrix. Experience has shown that GST, poly histidine, maltose binding protein and calmodulin binding protein are all good candidates for this approach. Obviously care must be taken in choosing the site at which affinity groups are inserted into the protein so that the tag is accessible and so that it does not disrupt function. This requires judicious choices coupled with trials of multiple sites and experimenting with different linker sequences.

If the affinity purification works very, very well the eluted proteins could be analyzed directly to determine both the intact masses of each component and in parallel the masses of their peptides for identification of the encoding genes. In this case elution of the complexes from the affinity matrix need not be gentle. However, to keep the background down and to deal with the likely possibility that the bait protein participates in more than one type of molecular complex, it will be advantageous to elute the complexes gently from the matrix so that the complexes remain in tact enabling further chromatographic fractionation. Perhaps the best elution method is to use a site-specific protease to cleave the tag sequence at an engineered site built into the construct in the linker region between the affinity tag and the protein sequence (Puig, et al, 2001; Rigaut, et al, 1999). (The remaining linker sequence could be engineered to have a second affinity sequence to facilitate recovery of the eluted products from the first column).

4.5 Subcontract or Consortium Arrangements

The majority of the work proposed in this Subproject will be performed at ORNL and PNNL. Additional research, as described above, will be performed at Sandia National Laboratories, Livermore, California; and the University of Utah. Letters of collaboration from lead investigators at these institutions are attached, and separate budgets are included from each institution.

4.6 Literature Cited

Back, J.W., Hartog, A.F., Dekker, H.L., Muijsers, A.O., de Koning, L.J., and de Jong, L. "A new crosslinker for mass spectrometric analysis of the quaternary structure of protein complexes." *J. Am. Soc. Mass Spectrom.* **12:** 222-227 (2001).

Baranov, P.V., Gurvich, O.L., Fayet, O., Prere, M.F., Miller, W.A., Gesteland, R.F., Atkins, J.F., and Giddings, M.C. Recode: A database of frameshifting, bypassing and codon redefinition utilized for gene expression. *Nuc. Acids Res* 264-267 (2001).

Belov, M.E., Anderson, G.A., Angell, N.H., Shen, Y., Tolic, N., Udseth, H.R., and 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**(5052-5060) (2001).

Belov, M.E., Gorshkov, M.V., Alving, K., and Smith, R.D. "Optimal Pressure Conditions for Unbiased External Ion Accumulation in a 2D RF-Quadrupole for FTICR Mass Spectrometry." *Rapid Commun. Mass Spectrom*. **15**: 1988-1996 (2001).

Belov, M.E., Gorshkov, M.V., Udseth, H.R., Anderson, G.A., and Smith, R.D. "Zeptomole-sensitivity Electrospray Ionization - Fourier Transform Ion Cyclotron Resonance." *Analytical Chemistry* **72**: 2271-2279 (2000).

Belov, M. E., Gorshkov, M. V., Udseth, H. R., Anderson, G. A., Tolmachev, A. V., Prior, D. C., Harkewicz, R., and Smith, R. D. "Initial Implementation of an Electrodynamic Ion Funnel with FTICR Mass Spectrometry." *Journal of the American Society for Mass Spectrometry* **11**: 19-23 (2000).

Bodanszky, M.; and Fagan, D. T. "Synthesis of Biocytin-Containing Peptides." *J. Am. Chem. Soc.* **99:** 235-239 (1977).

Boggs, P. T., Donaldson, J. R., Byrd, R. H., and Schnabel, R. B., Algorithm 676 ODRPACK: Software for Weighted Orthogonal Distance Regression. ACM Transactions on Mathematical Software (TOMS). New York, ACM Press. **15:** 348-364 (1989).

Caravatti, P. Method and Apparatus for the Accumulation of Ions in a Trap of an Ion Cyclotron Resonance Spectrometer, by Transferring the Kinetic Energy of the Motion Parallel to the Magnetic Field into Directions Perpendicular to the Magnetic Field, Spectrospin, AG (1988).

Cech, N. B. and Enke, C. G. "Relating Electrospray Ionization Response to Nonpolar Character of Small Peptides." *Analytical Chemistry* **72**: 2717-2723 (2000).

Cech, N. B. and Enke, C. G. "Effect of Affinity for Droplet Surfaces on the Fraction of Analyte Molecules Charged during Electrospray Droplet Fission." *Analytical Chemistry* 73: 4632-4639 (2001).

Cech, N. B., Krone, J. R., and Enke, C. G. "Predicting Electrospray Response from Chromatographic Retention Time." *Analytical Chemistry* **73**(2): 208-213 (2001).

Chen, T., Jaffe, J.D., and Church, G.M. "Algorithms for identifying protein cross-links via tandem mass spectrometry." *J. Comput. Biol* **8:** 571-583 (2001).

Cohen, F.E., and Sternberg, M.J.E. "On the use of chemically derived distance constraints in the prediction of protein structure with myoglobin as an example." *J. Mol. Biol.* **137:** 9-22 (1980).

Conrads, T.P., Alving, K., Veenstra, T.D., Belov, M.E., Anderson, G.A., Anderson, D.J., Lipton, M.S., Pasa-Tolic, L., Udseth, H.R., Chrisler, W.B., Thrall, B.D., and Smith, R.D. "Quantitative Analysis of Bacterial and Mammalian Proteomes Using a Combination of Cysteine Affinity Tags and 15N-Metabolic Labeling." *Analytical Chemistry* **73**: 2132-2139 (2001).

Conrads, T.P., Anderson, G.A., Veenstra, T.D., Pasa-Tolic, L., and Smith, R.D. "Utility of Accurate Mass Tags for Proteome-wide Protein Identification." *Anaytical Chemistry* **72**: 3349-3354 (2000).

de Wildt, R.M.T., Mundy, C.R., Gorick, B.D., and Tomlinson, I.M. "Antibody arrays for high-throughput screening of antibody-antigen interactions." *Nature Biotech.* **18:** 989-994 (2000).

Eng, J.K., McCormack, A.L., and Yates, J.R.I. "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database." *J. Am. Soc. Mass Spectrom.* **5:** 976-989 (1994).

Fields, S., and Song, O.K. "A novel genetic system to detect protein-protein interactions." *Nature* **340:** 245-246 (1989).

Flora, J.W., Hannis, J. C., and Muddiman, D. C. "High-Mass Accuracy of Product Ions Produced by SORI-CID Using a Dual Electrospray Ionization Source Coupled with FTICR Mass Spectrometry." *Analytical Chemistry* **73**: 1247-1251 (2001).

Gale, D. C. and Smith, R. D. "Small Volume and Low Flow-Rate Electrospray Ionization Mass Spectrometry of Aqueous Samples." *Rapid Communications in Mass Spectrometry* **7**(11): 1017-1021 (1993).

Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., and Cruciat, C.M., et al. "Functional organization of the yeast proteome by systematic analysis of protein complexes." *Nature* **415:** 141-147 (2002).

Gesteland, R.F. and Atkins, J.F. Recoding: Dynamic reprogramming of translation. In Annual Review of Biochemistry. C. C. Richardson, J. Abelson, A. Meister, and C.T. Walsh, eds. *Annual Reviews, Inc.,* Vol. 65, pp. 741-768 (1996).

Goodlett, D.R., Wahl, J.H., Udseth, H.R., and Smith, R.D. "Reduced Elution Speed Detection for Capillary Electrophoresis Mass-Spectrometry." *Journal of microcolumn separations* **5**(1): 57-62 (1993).

Griffin, T.J., Gygi, S.P., Rist, B., Aebersold, R., Loboda, A., Jilkine, A., Ens, W., and Standing, K.G. "Quantitative proteomic analysis using a MALDI quadrupole time-of-flight mass spectrometer." *Analytical Chemistry* **73:** 978-986 (2001).

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags." *Nature Biotech.* **17:** 994-999 (1999).

Hannis, J. and Muddiman, D. "A Dual Electrospray Ionization Source Combined with Hexapole Accumulation to Achieve High Mass Accuracy of Biopolymers in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry." *J. Amer. Soc. Mass Spectrom.* **11**: 876-883 (2000).

Havel, T.F., Crippen, G.M., and Kuntz, I.D. "Effects of distance constraints on macromolecular conformation. II. Simulation of experimental results and theoretical predictions." *Biopolymers* **18:** 73-81 (1979).

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., and Boutilier, K., et al. "Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry." *Nature* **415:** 180-183 (2002).

Jensen, P. K., Pasa Tolic, L., Anderson, G. A., Horner, J. A., Lipton, M. S., Bruce, J. E., and Smith, R. D. "Probing Proteomes Using Capillary Isoelectric Focusing-Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry." *Analytical Chemistry* 7**1**: 2076-2084 (1999).

Kim, T., Tang, K., Udseth, H. R., and Smith, R.D. "A Multi-Capillary Inlet Jet Disruption Electrodynamic Ion Funnel Interface for Improved Sensitivity Using Atmospheric Pressure Ion Sources." *Analytical Chemistry* **73**: 4162-4170 (2001).

Kornienko, O., Reilly, P.T.A., Whitten, W.B., and Ramsey, J.M. "Electron impact ionization in a microion trap mass spectrometer." *Rev. Sci. Instrum.* **70:** 3907-3909 (1999).

Lacroix, M., Rossi, V., Gaboriaud, C., Chevallier, S., Jaquinod, M., Thielens, N.M., Gagnon, J., and Arlaud, G.J. "Structure and assembly of the catalytic region of human complement protease C1r: A three-dimensional model based on chemical cross-linking and homology modeling." *Biochem.* **36:** 6270-6282 (1997).

Lee, S.-W., Berger, S. J., Martinovic, 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/FTICR." *Proc. Nat. Acad. Sci.* (2001).

Lehninger, A.L. "Biochemistry." **2nd edition:** New York: Worth. (1975).

Li, L., Masselon, C., Anderson, G.A., Pasa-Tolic, L., Lee, S.-W., Shen, Y., Zhao, R., Lipton, M.S., Conrads, T.P., Tolic, N., and Smith, R.D. "High-Throughput Peptide Identification from Protein Digests Using Data-Dependent Multiplexed Tandem FTICR Mass Spectrometry Coupled with Capillary Liquid Chromatography." *Analytical Chemistry* **73**(14): 3312-3322 (2001).

Lin, Y., and Matson, D. W. et al. Microfluidic Devices on Polymer Substrates for Bioanalytical Applications. Proceedings of the 3rd International Conference on Microreaction Technology, Frankfurt, Germany, Springer-Verlag (1999).

Lin, Y., and Xu, N. et al. Microfabricated dual-microdialysis and capillary isoelectric focusing devices for cleanup and separations/mass spectrometric analysis of biomolecules. Micro Total Analysis Systems 98. D. J. Harrison and A. V. Den Berg, Kluwer Academic Publishers**:** 343-346 (1998).

Loo, J.A. "Studying noncovalent protein complexes by electrospray ionization mass spectrometry." *Mass Spectrom. Rev.* **16:** 1-23 (1997).

Loo, J.A. "Electrospray ionization mass spectrometry: A technology for studying noncovalent macromolecular complexes." *Int. J. Mass Spectrom.* **200:** 175-186 (2000).

MacBeath, G., and Schreiber, S.L. "Printing proteins as microarrays for high-throughput function determination." *Science* **289:** 1760-1763 (2000).

Mann, M., and Wilm, M. "Error-tolerant identification of peptides in sequence databases by peptide sequence tags." *Analytical Chemistry* **66:** 4390-4399 (1994).

Marshall, A. G. "Milestones in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Technique Development." *International Journal of Mass Spectrometry* **200**(1-3): 331-356 (2000).

Marshall, A. G. and Guan, S. H. "Advantages of high magnetic field for Fourier transform ion cyclotron resonance mass spectrometry." *Rapid Communications in Mass Spectrometry* **10**(14): 1819-1823 (1996).

Martinovic, S., Berger, S.J., Pasa-Tolic, L., and Smith, R.D. "Separation and Detection of Intact Non-Covalent Protein Complexes from Mixtures by on-line Capillary Isoelectric Focusing-Mass Spectrometry." *Analytical Chemistry* **72**: 5356-5369 (2000).

Masselon, C., Tolmachev, A.V., Anderson, G.A., Harkewicz, R., and Smith, R.D. "Mass Measurement Errors Caused by Local Frequency Perturbations in FTICR Mass Spectrometry." *J. Amer. Soc. Mass Spectrom.* **13**: 99-106 (2002).

Mattson, G., Conklin, E., Desai, S., Nielander, G., Savage, M.D., and Morgensen, S. "A practical approach to crosslinking." *Mol. Biol. Rep.* **17:** 167-183 (1993).

Miranker, A.D. "Protein complexes and analysis of their assembly by mass spectrometry." *Curr. Opin. Struct. Biol.* **10:** 601-606 (2000).

Moore, B., Nelson, C., Giddings, M., Holmes, M., Kimball, M., Wills, N., Atkins, J. and Gesteland, R. DOE Genome Abstracts #33 (2002).

Mortz, E., O'Connor, P.B., Roepstorff, P., Kelleher, N.L., Wood, T.D., McLafferty, F.W., and Mann, M. "Sequence tag identification of intact proteins by matching tandem mass spectral data against sequence databases." *Proc. National Acad. Sci. USA* **93:** 8264-8267 (1996).

Muller, D. R., Schindler, P., Towbin, H., Wirth, U., Voshol, H., Hoving, S., and Steinmetz, M. O. "Isotope-Tagged Cross-Linking Reagents. A New Tool in Mass Spectrometric Protein Interaction Analysis." *Analytical Chemistry* **73:** 1927-1934 (2001).

Pasa-Tolic, L., Harkewicz, R., Anderson, G.A., Tolic, N., Shen, Y., Zhao, R., Thrall, B.D., Masselon, C., and Smith, R.D. (Submitted). "Increased Proteome Coverage Based upon High Performance Separations and DREAMS FTICR Mass Spectrometry." *Analytical Chemistry* **submitted**.

Pasa-Tolic, L., Jensen, P. K. , Anderson, G. A. , Lipton, M. S. , Peden, K.K. , Martinovic, S. , Tolic, N. , Bruce, J. E. and Smith, R. D. "High Throughput Proteome-Wide Precision Measurements of Protein Expression using Mass Spectrometry." *J. Amer. Chem Soc.* **121**: 7949-7950 (1999).

Pearson, K.M., Pannell, L.K., and Fales, H.M. "Intramolecular cross-linking experiments on cytochrome c and ribonuclease A using an isotope multiplet method." *Rapid Commun. Mass Spectrom.* **16:** 149-159 (2002).

Puig, O., Casperry, F., Rigaut, C., Rutz, B., Bouveret, E., Bradago-Nilsson, E., Wilm, M. and Seraphin, B. *Methods* **24**, 218-229 (2001).

Rappsilber, J., Siniossoglou, S., Hurt, E.C., and Mann, M. "A generic strategy to analyze the spatial organization of multi-protein complexes by crosslinking and mass spectrometry." *Analytical Chemistry* **72:** 267-275 (2000).

Reilly, P.T.A., Kornienko, O., Whitten, W.B., and Ramsey, J.M. "Microscale ion traps: 2-dimensional arrays." *Proceedings of the 48th ASMS Conference on Mass Spectrometry and Allied Topics,* Long Beach, CA, 635, (2000).

Reineke, U., Volkmer-Enger, R., and Schneider-Mergener, J. "Applications of peptide arrays prepared by the SPOT-technology." *Curr. Opin. Biotech.* **12:** 59-64 (2001).

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. "A generic protein purification method for protein complex characterization and proteome exploration." *Nature Biotech.* **17:** 1030-1032 (1999).

Rigaut, C., Shevchjenko, A., Rutz, B., Wilm, M., Mann, M. and Sersphin, B. *Nature Biotechnology* **17**,1030-1032 (1999).

Rodi, D.J., and Makowski, L. "Phage-display technology - finding a needle in a vast molecular haystack." *Curr. Opin. Biotech.* **10:** 87-93 (1999).

Rumelhart, D. and McClelland, J. "Parallel Distributed Processing: Explorations in the Microstructure of Cognition." MIT Press (1986).

Shah, A.A., Giddings, M.C., Parvaz, J.B., Gesteland, R.F., Atkins, J.F., and Ivanov, I.P. Computational identification of putative programmed translational frameshift sites in protein-encoding nucleotide sequences. (submitted to Bioinformatics on 8/8/01) (2002).

Shaffer, S. A., and Prior, D. C. et al. "An Ion Funnel Interface For Improved Ion Focusing and Sensitivity Using Electrospray Ionization Mass Spectrometry." *Analytical Chemistry* **70**(19): 4111-4119 (1998).

Shaffer, S. A., Tang, K. Q., Anderson, G. A., Prior, D. C., Udseth, H. R., and Smith, R. D. "A Novel Ion Funnel For Focusing Ions At Elevated Pressure Using Electrospray Ionization Mass Spectrometry." *Rapid Communications in Mass Spectrometry* **11**(16): 1813-1817 (1997).

Shen, Y., and Tolic, N. et al. "High-Throughput Proteomics using High Efficiency Multiple-Capillary Liquid Chromatography with on-line High Performance ESI FTICR Mass Spectrometry." *Analytical Chemistry* **73**(13): 3011-3021 (2001).

Shen, Y., Xiang, F., Veenstra, T.D., Fung, E.N., and Smith, R.D. "High-Resolution Capillary Isoelectric Focusing of Complex Protein Mixtures from Lysates of Microorganisms." *Analytical Chemistry* **71**: 5348-5353 (1999).

Shen, Y., Zhao, R., Belov, M.E., Conrads, T.P., Anderson, G.A., Tang, K., Pasa-Tolic, L., Veenstra, T.D., Lipton, M.S., Udseth, H.R., and Smith, R.D. "Packed Capillary Reversed-Phase Liquid Chromotaography with High-Performance Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Proteomics." *Analytical Chemistry* **73**: 1766-1775 (2001).

Shulttz, G. A, Corso, T. N., and Prosser, S. J. "Proceedings 49th ASMS Conference on Mass Spectrometry and Allied Topics." (2001).

Smith, R.D., Anderson, G.A., Lipton, M.S., Pasa-Tolic, L., Shen, Y., Conrads, T.P., Veenstra, T.D., and Udseth, H.R. "An Accurate Mass Tag Strategy for Quantitative and High Throughput Proteome Measurements." *Proteomics:* In press (2002).

Smith, R.D., Bruce, J.E., Wu, Q.Y., and Lei, Q.P. "New mass spectrometric methods for the study of noncovalent associations of biopolymers." *Chem. Soc. Rev.* **26:** 191-202 (1997).

Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J., and Udseth, H. R. "New Developments in Biochemical Mass Spectrometry: Electrospray Ionization." *Analytical Chemistry* **62**(9): 882-899 (1990).

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." *Electrophoresis* **22**: 1652-1668 (2001).

Subramanian, A., Kennel, S.J., Oden, P.I., Jacobson, K.B., Woodward, J., and Doktycz, M.J. "Comparison of techniques for enzyme immobilization on silicon supports." *Enzyme and Microbial Tech.* **24:** 26-34 (1999).

Tang, K., Lin, Y., Matson, D.W., Kim, T., and Smith, R.D. "Generation of Multiple Electrosprays Using Microfabricated Emitter Arrays for Improved Mass Spectrometric Sensitivity." *Analytical Chemistry* **73**: 1766-1775 (2001).

Tolmachev, A.V., Kim, T., Udseth, H.R., Smith, R.D., Bailey, T.H., and Futrell, J.H. "Simulation-Based Optimization of the Electrodynamic Ion Funnel for High Sensitivity Electrospray Ionization-Mass Spectrometry." *Int. J. Mass Spec.* **2003**: 31-47 (2001).

Tolmachev, A.V., Masselon, C., Anderson, G.A., Udseth, H.R., and Smith, R.D. "Frequency Shifts Due to Interface of Resolved Peaks in Magnitude-Mode Fourier-Transform Ion Cyclotron Resonance Spectra." *J. Am. Soc. Mass Spectrom.* **13**(4): 387-401 (2002).

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., and Pochart, P., et al. "A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*." *Nature* **403:** 623-627 (2000).

Wahl, J.H., Goodlet, D.R., Udseth, H.R., and Smith, R.D. "Use of small-diameter capillaries for increasing peptide and protein detection sensitivity in capillary electrophoresis-mass spectrometry." *Journal of the American Chemical Society* **115**: 803-804 (1993).

Wallon, G., Rappsilber, J., Mann, M., and Serrano, L. "Model for stathmin/OP18 binding to tubulin." *EMBO J.* **19:** 213-222 (2000).

Washburn, M.P., Wolters, D., and Yates, J.R. "Large-scale analysis of the yeast proteome by multidimensional protein identification technology." *Nature Biotech.* **19:** 242-247 (2001).

Wen, J., Lin, Y., Xiang, F., Matson, D.W., Udseth, H.R., and Smith, R.D. "Microfabricated Isolelectric Focusing Device for Direct Electrospray Ionization Mass Spectrometry." *Electrophoresis* **21**: 191-197 (2000).

White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D. L., Moffat, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M. Vamathevan, J. J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M. J., Minton, K.W., Fleischmann, R. D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J., Craig, F., and Claire, M. "Genome Sequence of the Radioresistant Bacterium Deinococcus radiodurans R1." *Science* **286**: 1571-1577 (1999).

Wilkinson, J. Fragmentation of Polypeptides by Enzymic Methods. Practical Protein Chemistry–A Handbook. A. Darbre, John Wiley & Sons (1986).

Wilm, M. S. and Mann, M. "Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last?" *International Journal of Mass Spectrometry and Ion Processes* **136**(2-3): 167-180 (1994).

Wolters, D.A., Washburn, M.P., and Yates, J.R. "An automated multidimensional protein identification technology for shotgun proteomics." *Analytical Chemistry* **73:** 5683-5690 (2001).

Wong, S.S. "Chemistry of Protein Conjugation and Crosslinking." Boca Raton: CRC Press, Inc. (1991).

Yao, X.D., Freas, A., Ramirez, J., Demirev, P.A., and Fenselau, C. "Proteolytic O-18 labeling for comparative proteomics: Model studies with two serotypes of adenovirus." *Analytical Chemistry* **73**: 2836-2842 (2001).

Young, M.M., Tang, N., Hempel, J.C., Oshiro, C.M., Taylor, E.W., Kuntz, I.D., Gibson, B.W., and Dollinger, G. "High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry." *Proc. National Acad. Sci. USA* **97:** 5802-5806 (2000).

Zhu, H., Klemic, J.F., Chang, S., Bertone, P., Casamayor, A., Klemic, K.G., Smith, D., Gerstein, M., Reed, M.A., and Snyder, M. "Analysis of yeast protein kinases using protein chips." *Nat. Genet.* **26:** 283-289 (2000).

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., and Houfek, T., et al. "Global analysis of protein activities using proteome chips." *Science* **293:** 2101-2105 (2001).

4.8 Milestones

Tasks	FY03	FY04	FY05
4.4.1.1 Increase sensitivity of FTICR mass measurements	X	X	
4.4.1.2 Increase the dynamic range of protein complex measurements	X	X	
4.4.1.3 Explore an alternative AMT based approach using Q-TOF technology	X	X	
4.4.2.1 Develop capabilities for MS analysis of crosslinked proteins in complexes	X	X	
4.4.2.2 Protein Chip/Ion Trap Mass Spectrometer Array Read Out System		X	X
4.4.3.1 Predict peptide capillary LC elution times to aid in protein identification	X	X	
4.4.3.2 Increase the FTICR mass measurement accuracy using AMTs		X	X
4.4.3.3 Improved quantitation and measures of data quality		X	
4.4.4.1 Linked LC-MS and LC-FTICRMS	X	X	
4.4.4.2 New chromatographic-MS techniques for identifying protein-protein interactions	X	X	
4.4.4.3 Global approaches for protein identification by LC-MS	X	X	X

This page intentionally left blank.

Appendix

This page intentionally left blank.