# 3.0 HIGH THROUGHPUT COMPLEX PROCESSING

**Abstract**

In order to achieve the stated Genomes to Life Goal of identifying 80% of the protein machines in a given organism within one year, it will be necessary to develop a technological approach capable of high throughput purification of protein complexes at nanoscale volumes, combined with a highly efficient on-line coupling to mass spectrometry-based protein identification, and with minimal genetic manipulation of the organism of interest. This project is focused on developing an integrated technological approach to this problem, based on the availability of a fully sequenced genome and associated open reading frames. This provides the foundation for generation of recombinant antibodies capable of immunoprecipitating protein complexes. Renewable surface microcolumns are used for complex isolation, purification, and concentration. Microfluidic devices are included in the system to provide in-line separation of proteins and peptides, and to provide an alternative mechanism for capture of protein complexes from small batches of cells. Each of these components has distinct advantages for maximizing high throughput and sensitivity while emphasizing physiologically relevant complexes. Back-up strategies to each of these components have been discussed in the Core project. The optimization and integration of these components for high throughput sample processing will be developed in the following specific aims:

Specific Aim #1: To use high throughput selection of single chain variable fragment (scFv) antibodies in yeast to generate a comprehensive library of high affinity scFv antibodies suitable for isolation of protein complexes.

Specific Aim #2: To develop a renewable surface microcolumn system compatible with isolation of protein complexes using immobilized scFv or other affinity matrixes and subsequent on-line digestion to generate peptides for mass spectrometric analysis.

Specific Aim #3: To develop microfluidic devices for integrated processing and analysis of protein complexes in conjunction with on-line mass spectrometry.

**ORNL:**
J. Michael Ramsey, Ph.D.
Chemical Sciences Division
Oak Ridge National Laboratory
P.O. Box 2008, MS-6142
Oak Ridge, TN 37831-6142
Phone: 865/574-5662
Facsimile: 865/574-8363
E-mail: ramseyjm@ornl.gov

**PNNL:**
Karin D. Rodland, Ph.D.
Molecular Biosciences Division
Pacific Northwest National Laboratory
P.O. Box 999/MS P7-56
Richland, WA 99352
Phone: 509/376-7605
Facsimile: 509/509-6767
E-mail: Karin.Rodland@pnl.gov

**This page intentionally left blank.**

## 3. 1 Budget Pages and Budget Explanation

## 3.2 Background and Significance

We propose to develop automated, high throughput methods to isolate protein complexes, and for subsequent identification and characterization. One approach to be taken will be to produce single chain variable fragment (scFv) antibodies specific to protein complexes of interest. These antibodies will be incorporated into renewable surface microcolumn systems for affinity separations of the proteins and complexes. Affinity-based separations provide highly selective isolation of the proteins and complexes from their matrices and also concentrate the samples prior to subsequent processing or analysis. After separation, the isolated complexes can then be analyzed as intact proteins or can be enzymatically digested on-line to generate unique peptides. Initially, this work will be executed using conventional technology while automated, high throughput platforms, such as microfluidic devices, are being developed in parallel. Using microfabricated architectures for high throughput processing will enable serial cou0ling of all fluid handling and processing steps under computer control without operator intervention under computer control. In addition to serial integration of the processing steps, the microfluidic architecture can be easily replicated to process smaples in parallel. With successful development of the high throughput processing, proteins, protein complexes, and their peptides from single cells and ensembles of cells can be processed and analyzed rapidly. The background, rationale and research for each aspect are described in detail below.

### 3.2.1 Current approaches for isolation of protein complexes

The study of protein complexes obviously requires a method for capturing biologically relevant protein complexes that retain their native abilities for interaction while simultaneously removing extraneous proteins. Three basic approaches have been utilized in most studies: (1) direct isolation, (2) genetic 'two-hybrid' screens, and (3) affinity separation approaches. Each of these approaches has distinct fundamental weaknesses and strengths.

**Direct isolation:** Direct isolation requires some biochemical property than can be exploited for purification, such as size or DNA binding ability, as well as a functional assay to assess purification. This approach is limited to well defined complexes such as ribosomes or holoenzymes with DNA or RNA polymerase activity. Direct isolation approaches can be improved by the use of chemical cross-linking reagents to establish robust covalent linkages that can be subjected to rigorous biochemical purification procedures. However, there is always concern as to whether the results of a cross-linking reaction represent specific interacting proteins or merely reflect proteins in close proximity. Accessibility of the cross-linking reagent to the cellular compartment of interest can also be problematic.

**Yeast two-hybrid screen:** The yeast two-hybrid screen illustrates the classic advantages and disadvantages of the genetic approach. Although the ability of the yeast two hybrid approach to screen a large library of clones for potential binding partners is a distinct advantage and avoids *a priori* assumptions about likely partners, the method is limited to screening genetically engineered 'bait proteins' one at a time, and each bait protein requires a separate cloning and transduction. This approach is limited to identifying two or at most three components of a given complex; if full activation requires more than two or three proteins, the accessory members will not be identified by this strategy. Furthermore, the assay used for selection is highly artificial and unrelated to the probable biological function of the native complex in the parent organism. Biologically important complexes may fail to form in yeast because an essential intermediary protein is not expressed.

**Affinity-based purification:** Co-purification strategies based on affinity reactions with one specific member of the complex have the advantage of being able to isolate complexes that were formed under relevant physiological conditions inside the organisms or cell type of interest. Multi-protein complexes can be purified, and each individual component identified. Two basic approaches are available: *in vivo* expression of affinity tagged bait proteins and *in vitro* displacement using affinity tagged bait proteins. The "affinity reagent" is a fusion protein that consist of the bait protein fused to a well known affinity tag such as GST, 6His, HA, myc etc. In genetically tractable organisms the affinity tagged bait protein can be expressed *in vivo*. The complexes form *in vivo*, and are

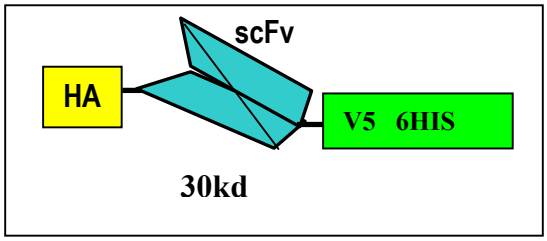
then purified from cell lysates on the appropriate affinity matrix by using the affinity handle that is on the bait. This method can only be used with genetically tractable organisms and some organisms that will be examined in the "Genomes to Life program" are genetically intractable, thus requiring an alternate approach.

One proposed alternative approach avoids the necessity of expressing the affinity tagged protein in the organism of interest. Instead, the purified affinity tagged bait is added to a cellular lysate of interest. Equilibrium kinetics and counterpart displacement allow the affinity tagged bait protein to displace its counterpart in the complex within the lysate. Then, as above, complexes form, and the complexes are purified from the cell lysates on the appropriate affinity matrix using the affinity handle that is on the bait. A strength of this approach is that a single lysate can be screened with many different purified affinity tagged baits. However, the generation of affinity reagents is the rate-limiting step in this process. Both of these approaches suffer from a number of concerns, including the possibility that overexpression of the bait protein might drive reactions that would not occur at lower protein levels. This could lead to the detection of many artificial interactions. In addition, the genetic modification to the bait protein may interfere with some physiologically relevant interactions, thus precluding the formation of native complexes.

**Immunoprecipitation:** Immunoprecipitation of intact complexes with high affinity antibodies has the advantage of targeting complexes formed under physiological conditions, without altering the protein constituents of the complex. Immunoprecipitation is also readily adaptable to parallel analysis of multiple distinct complexes. However, the bottleneck lies in the need to generate high affinity antibodies for specific proteins of interest. Traditional approaches using mice and rabbits to generate antibodies require prior knowledge about the proteins of interest, and either purified protein or defined synthetic peptides to act as antigens. Antibody generation and characterization is time consuming and expensive. However, recombinant approaches such as the scFv method described in this proposal have the potential to substantially solve the problem of antibody generation.

The scFv technology uses genetic engineering to combine the antigen-specific variable domains from both the heavy ($V_H$) and light ($V_L$) chains of immunoglobulin molecules into a single DNA coding sequence. This DNA, when placed in suitable organism, directs the production of a single-chain variable fragment (scFv) that has the ability to bind to a portion of a protein**.** This svFV technology is highly compatible with the equlibrium kinetics and counterpart displacement approach discussed above. Since the displacement approach requires the purification of recombinant bait proteins, we have at our disposal a ready source of antigen with which we can rapidly generate multiple scFv for different epitopes on the bait protein. The multiple scFv to the different epitopes allow for a tandem affinity purification of the complexes. Considering the depth at which we will be mining the bacterial proteomes, we expect to generate scFv for several proteins in each complex. The multiple entry point approach to purifying the complexes helps validate the specificity of components attributed to a given complex. For example, non-specific interactions are less likely to form on multiple entry points. Furthermore, the secreted form of the scFv is constructed as a fusion protein with three affinity tags HA, V5, and 6HIS, that are used to purify the complex (**Fig. 3.2.1**). The redundancy of the purifications and multiple entry points will help validate the true components of the complexes.

**Fig. 3.2.1:** Secreted form of scFv with three affinity




The scFv approach has a number of strengths over the before mentioned approaches. The scFv is the only approach that allows purification of native complexes. The complex components are under normal physiologic expression patterns and levels, thereby reducing the false positives inherent in the alternative techniques. The scFv also provides a tool that can used to further investigate the complex, for example, in immunoblots, cellular localization, and transdominant inhibition studies.

Although the scFv technology in this format is fairly new, scFv have been pursued for over 15 years. ScFv utility has been demonstrated in a number of approaches; ELISA, immunoblotting, cellular staining, and immuno-precipitation of a complex followed by mass spectrometry.

We were the first to utilize scFv antibodies in this capacity [Siegel et al., 2002]. A scFv antibody, expressed in a monomeric soluble form recognizing human Ku86 protein, was used for immunoprecipitation reactions with HeLa cell extract and the affinity-isolated precipitants were analyzed by MALDI-TOF mass spectrometry following enzymatic digestion. Not only was the target antigen (Ku86) identified in these experiments, but its heterodimeric binding partner, Ku70, was as well. We plan to build upon our expertise to extend this approach to a high throughput scale by coupling the scFv antibodies with the renewable surface matirx format discussed below.

### 3.2.2 Development of a high throughput approach to analysis of protein complexes

An optimal high throughput approach to the identification of protein complexes would provide parallel analyses of a single biological sample, without the need to genetically modify either the host organism or the proteins of interest. The biochemical purification procedure would be amenable to on-line sample evaluation and detection, and could be accomplished with nanoscale technologies. The increase in data output derived from parallel sample analysis would permit statistical approaches to data interpretation, including more efficient discrimination between specific and non-specific interactions.

We are proposing an integrated sample processing scheme (illustrated in **Fig. 3.2.2**) in which protein complexes will be isolated using affinity-based approaches, either through recombinant affinity tags and appropriate matrices, or by immunoprecipitation with recombinant scFv antibodies selected for high affinity binding. Affinity-purified protein complexes will be captured on renewable surface microcolumns, which can be arrayed to optimize parallel processing of multiple complexes obtained under a single set of experimental conditions. The

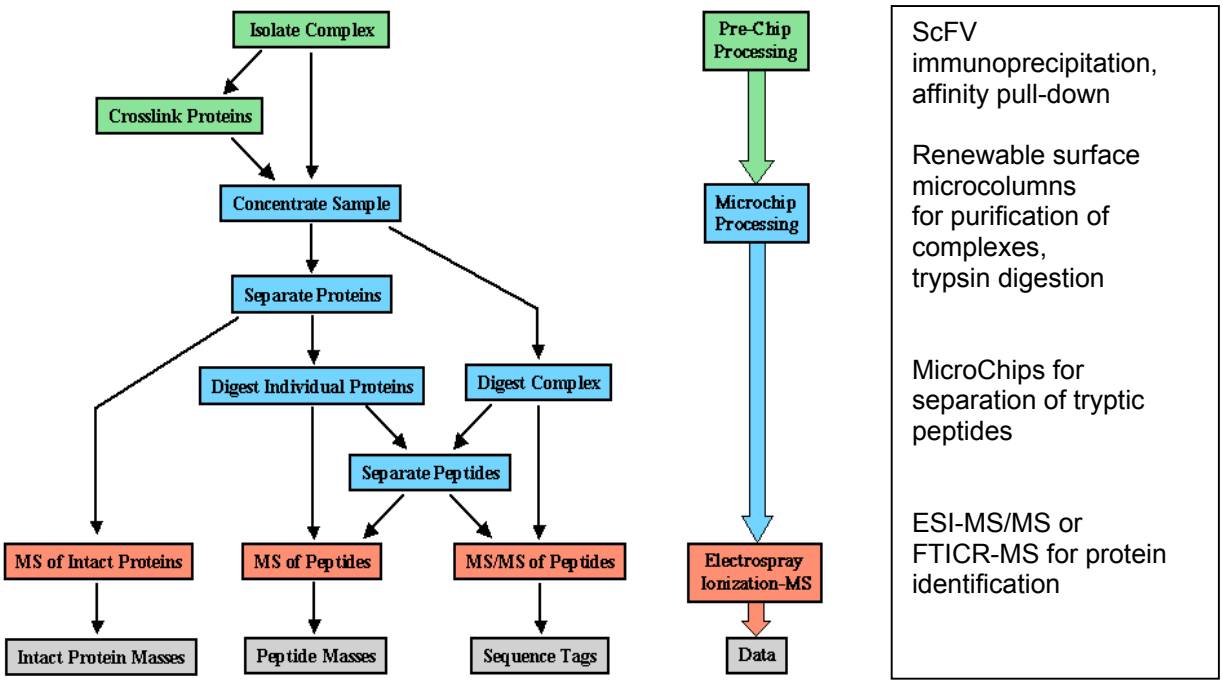



**Fig. 3.2.2.** Flow chart for high throughput processing

renewable surface microcolumns also provide a suitable surface for efficient trypsin digestion of the protein complexes. The resulting tryptic peptides will be sent to microfluidic devices for two dimensional separation based on C18 reverse-phase open-channel electrochromatography (OCEC) and capillary electrophoresis (CE) followed by on-line mass spectrometric analysis.

#### 3.2.2.1 Development of scFv antibodies for multiplexed purification of protein complexes

Immunoprecipitation of protein complexes is highly amenable to parallel processing of a single biological sample; the rate limiting step to complex isolation is simply the availability of a sufficient diversity of antibodies with adequate specificity and affinity. The major bottleneck to this approach in the past has simply been the availability of appropriate antibodies. Antibodies have traditionally been developed in a 'one off' fashion, to specific proteins of defined functional interest, and problems of cross-reactivity and modest affinity have contributed to the technical difficulties of using immunoprecipitation to efficiently isolate protein complexes. We propose to use a multiplexed recombinant scFv system which can potentially generate antibodies to any theoretical protein while simultaneously providing a mechanism for affinity selection that minimizes problems associated with traditional polyclonal or monoclonal antibodies generated against synthetic peptides [Boder et al., 1998; 2000].

Phage display systems have circumvented many of the traditional problems associated with generation of antibodies *in vivo* by using *in vitro* methods to select antibody fragments derived from human B cells (see Belov et al., 2001). This technology links the antigen-specific immunoglobulin variable domains from both the heavy ($V_H$) and light ($V_L$) chains into a single DNA coding sequence allowing the resulting single-chain variable fragment (scFv) antibody to be isolated after binding to an antigen of interest. Although phage libraries have been used to isolate antibodies against many antigens, the procedure is still not widely used due, in part, to the technical expertise required to create high quality libraries of sufficient diversity (necessary for high affinity antibodies) and difficulties associated with amplifying those libraries for subsequent experiments (limiting the number of screens which can be performed before recreating the library). [Holt et al., 2000]

Surface display of scFv libraries on yeast circumvents the problems associated with library propagation found in phage. ScFv libraries are easily expanded since the scFv's are not expressed under the growth conditions used during library amplification. The ability to completely inhibit expression of the scFv antibody during library amplification provides conditions where all clones can be represented equally as the growth of each clone is unaffected by the unwanted expression of antibodies that may induce a slow growth phenotype. Using such a tightly regulated expression system, not currently available for phage display, eliminates the loss of complexity and normalization in the library due to the over representation of clones that are more easily expressed. Therefore after a library is generated, characterized, and validated, it can be used many times to identify an almost unlimited number of scFvs for tens of thousands of antigens. Potentially, all scFv technologies can produce antigen-specific scFv that can be used for a multitude of applications such as immunoprecipitation, ELISA, immunohistochemistry, flow cytometry, protein arrays, and therapeutics.

The scFv technology can provide reagents for literally any protein. Standard monoclonal/polyclonal antibody technologies are frequently unsuitable for generating antibodies to specific epitopes and proteins because of the conserved nature of proteins between species and self-reactivity. Our approach will achieve comparable result at a greatly reduced development time line and cost. Regardless of the protein or pathway being investigated, a source of high affinity, highly specific, stable, and easily produced antigen-specific reagents that are amenable to standard surface chemistries would be a valuable resource for many areas of proteomics research today and in the future.

The ability to generate extensive libraries with binding affinities in the nM range provides a key set of reagents for protein purification, isolation of protein complexes, and functional studies of specific proteins. Furthermore, the recombinant antibodies generated by our scFv strategy in yeast can be readily engineered to contain affinity tags which provide novel mechanisms for immobilization and subsequent purification and characterization. In particular, scFv antibodies can be coupled with renewable surface systems for complex isolation described below, to provide an integrated high throughput system.

### 3.2.3 Renewable surface isolation of complexes for mass spectrometry

Renewable surface (RS) techniques can be used to rapidly bind protein complexes from cell extracts onto microbeads derivatized with specific binding sites (e.g., single chain antibodies, with simultaneous observation of

protein binding events using an optical biomolecular interaction assay (BIA) method. When a specific binding event is observed, the proteins will be eluted, further processed, and identified using mass spectrometry. In this way, only positive binding results will be further analyzed, therefore minimizing the accumulation of mass spectrometry data. In addition, the protein binding/elution information obtained by on-column optical detection can be used to optimize the protein binding/elution conditions, and control the delivery of reagents for downstream processing. Mass spectrometry is ideal for identifying enzymatically digested proteins using liquid chromatography (LC) followed by tandem mass spectrometry (MS/MS) for initial identification, and high sensitivity, high resolution fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) once accurate mass tags are identified [Conrads et al., 2000; Lipton et al., 2002; Smith et al., 2002]. The combination of automated renewable surface methods, biomolecular interaction assay, and mass spectrometry techniques (RS-BIA-MS) will therefore provide rapid analysis of protein complexes. PNNL is a leader in both the development of mass spectrometry methods for protein identification [ Conrads et al., 2000; 2001; Lipton et al., 2002; Smith et al., 2001; 2002; Belov et al., 2001; Li et al., 2001; Shen et al., 2001] and the development of renewable surface systems for bioseparation and biosensing [Bruckner-Lea et al., 1998; 1999; 2000; 2001], so we are in a unique position to combine these techniques for proteomic analysis.

#### 3.2.3.1 Description of the renewable surface approach for biomolecular interaction assays

In the renewable surface approach for biomolecular interaction assays, a suspension of surface derivatized beads is introduced into a flow system and then automatically trapped by a barrier that stops the beads but allows the fluid to proceed. This produces a small microcolumn of beads (only about 1 microliter in volume) that can be automatically perfused with reagent or sample solutions to perform separations or surface reactions. Simultaneous measurements of absorbance or fluorescence via optical fibers incorporated into the flow cell are used to detect protein binding. At the completion of the observation, the beads can be automatically flushed from the observation area and delivered downstream for further analysis or disposal. A new bed of beads is then automatically packed for the next assay. This new bead bed has a fresh surface, hence the name "renewable surface" sensing. The RS-BIA technique was first demonstrated by Professor Ruzicka at the University of Washington [Ruzicka, 1998; Ruzicka & Scampavia, 1999; Ruzicka & Hansen, 2000] and we have recently expanded upon the technique and developed new flow cells (two patents [Egorov et al., 1999; Grate et al., 1999] and two patents currently pending).

The delivery and manipulation of reagents, samples, wash solutions and bead suspensions for RS-BIA is carried out using a computer-operated sequential injection (SI) system. Sequential injection is a recent variant of flow injection methods that provides greater precision and versatility while being more compatible with computer control. This approach is an open architecture fluid handling method giving the experimentalist complete control over system design, materials, and operation. It can be mated with autosamplers and fraction collectors for automating high throughput analyses The renewable surface biomolecular interaction assay is well-suited for the automated isolation of complexes for high throughput analysis, while providing the following advantages:

- Automated fluid handling.
- Affinity surfaces are automatically delivered and localized in the observation area for every protein capture experiment.
- Each affinity surface has never been previously used, contaminated or regenerated (identical surface chemistry for each analysis, no carryover).
- The column format is well suited for perfusion, enabling rapid sample processing (minutes).
- Ability to deliver captured proteins to a downstream detector for more detailed characterization.
- An optical signal indicates that proteins have been captured/eluted. The optical signal can be used to trigger downstream analysis, downstream reagent delivery, and optimize binding/elution conditions.
- Small volumes can be handled (10-1000 ul solution volumes are routinely handled. Column size is typically only 1-5 µl packed bed volume).
- Can deliver solution volumes and protein concentrations required for downstream mass spectrometry.
- Enables automated, reliable, high throughput processing.
- Multiple columns could be used in parallel for even higher throughput analysis.

### 3.2.4 Microfluidic devices for integrated processing and analysis of protein complexes

Conventional analysis of isolated protein complexes has traditionally involved separation of the constituent proteins by SDS-PAGE or 2D PAGE followed by extraction of the individual proteins for enzyme digestion and MS analysis of the peptides. Direct analysis of peptides from digests of whole protein complexes may also be performed by 1- or 2-dimensional HPLC separations with on-line MS or MS/MS. The latter approach has the potential to greatly increase analytical throughput rates by avoiding the gel electrophoresis and extraction steps. On-line separation prior to MS detection also allows removal of matrix components and provides temporal resolution of the peptides, which improve signal-to-noise ratios and result in simpler, easier to interpret mass spectra. Use of microfabricated separation devices, such as the microchip technology to be developed in this project, allows for even greater throughput rates. The time-scale of microchip separations is 1/10th to 1/100th that of conventional LC and will allow efficient use of the high-speed analysis capabilities of MS. On-line MS analysis has been achieved by direct microchip-MS coupling via capillary electrospray ionization tips, as discussed in more detail in Preliminary Data. Automated sample processing steps such as concentration, labeling and enzymatic digestion may be integrated on these devices to reduce manual operations and improve reproducibility. In addition to the time-scale advantages of microchip analysis, the physical dimensions of microchannels in the devices are particularly well-suited for manipulation and analysis of individual microbial cells . Rapid analysis of single human tumor cells by capillary electrophoresis (CE) with on-column cell lysis has been shown to give significantly different electropherograms from that of batch processed cells, presumably due to non-metabolic reactions which occur during the longer batch workup. The CE detection of proteins from single cells has been performed by the highly sensitive technique of laser-induced fluorescence. On-chip lysis and processing of small batches of cells could potentially provide sufficient complex for analysis by nanoelectrospray ionization mass spectrometry.

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 Center sites at both ORNL and PNNL.

## 3.3 Preliminary Data

### 3.3.1 Generation of an scFv antibody library

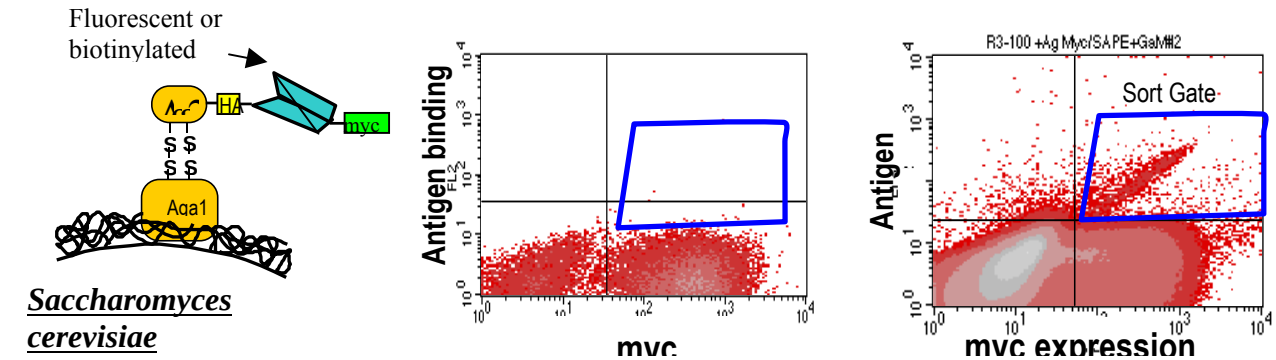
#### 3.3.1.1 Properties of scFv antibodies

Our collaborator Dane Wittrup at MIT has developed a general method for engineering the binding and stability properties of scFvs by expressing them on the surface of yeast cells (**Fig. 3.3.1A**; [Boder et al., 1998; 2000]. Yeast surface display presents two key advantages over other *in vitro* alternatives such as display on the surface of bacteriophage: 1) the kinetics and thermodynamics of antigen binding reactions can be measured quantitatively and used to design optimal screening strategies; and 2) yeast is a single-celled eucaryote and therefore possesses biosynthetic secretory apparatus analogous to that of B cells, allowing greater functional diversity of expressed scFvs compared to bacterial systems. The figures below illustrate typical results from flow cytometric analysis where the cells have been double-labeled as shown in **Fig. 3.3.1B** and **Fig. 3.3.1C** in order to independently detect antigen binding events and to normalize for random variation in cell-to-cell scFv expression levels. The antigen dissociation rate constant $k_{off}$ can be measured by competition with excess unlabeled antigen as shown in **Fig. 3.3.1E**, with precision ±10%, and the equilibrium binding constant $K_D$ can be measured as shown in **Fig. 3.3.1D** with precision ± 30%.

#### 3.3.1.2 Antibody library construction

**A library of 1 x $10^9$ human scFv antibody fragments has been constructed in the yeast display format:** Poly A mRNA from the lymph nodes and spleens of 57 anonymous adults were obtained from commercial sources and pooled (Origene, Clontech, Biochain). First strand cDNA synthesis was primed with IgM and IgG primers as

described previously (*JMB* 222:581, '91). V gene PCR primers designed to amplify all major human germline gene families were used to separately amplify VH and VL coding regions (*Immunotechnology* 3:271, '98). The diversity of the cDNA template was confirmed as > $10^7$ for each reaction by extrapolation to zero time with RT-PCR. Amplified VH and VL fragments were separately cloned (8 x $10^7$ VH, 1 x $10^6$ VL), then assembled together with a $(G_4S)_3$ linker as previously described (*PNAS* 95:6157, '98). The potential diversity of this recombined scFv library would be 8 x $10^{13}$. 90 assembled scFvs were sequenced, and the V gene usage was identified using the online VBASE database; the sequenced pool contained *6/7 functional VH gene families , 5/6 Vμ, and 6/10 of the Vμ genes* 1 x $10^9$ yeast transformants were obtained from this scFv pool, and the yeast



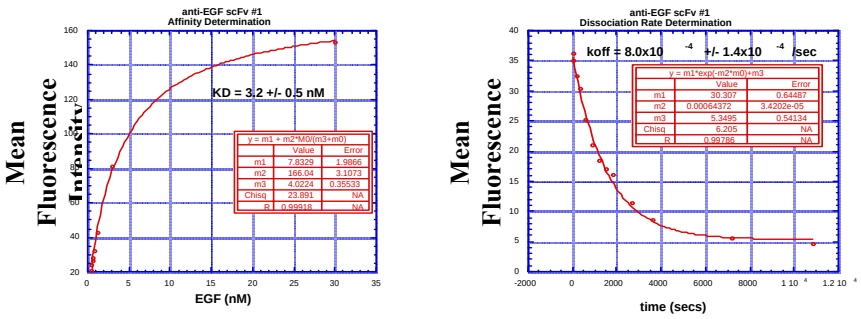




**Fig. 3.3.1**. Panel A depicts the expression construct of the scFv on the surface of yeast. Panel B & C, Flow cytometric sorting of EGF binding scFv clones from $10^9$ diverse scFv library. Panel D & E, solution based experiments analyzed by flow cytometry to determine the affinity ($K_D$) and dissociation rate (k off) of one scFv antibody recognizing human EGF isolated by yeast display. $K_D$ was determined by measuring fluorescence of labeled EGF at concentrations from 0.3 to 30 nM. The calculated affinity constant was 3 nM. The dissociation rate was determined by saturating the antibodies with labeled EGF and then measuring the fluorescence remaining after competition with 100-fold molar excess of unlabeled EGF at various time points. The off rate for this clone was 8x10-4/sec.

displayed library was amplified and frozen. 90 individual library clones were induced and scFv expression determined using the C-terminal c-myc epitope tag. Approximately 2/3 of the examined clones displayed scFv at greater than twice the background signal. *Therefore, the majority of the cloned repertoire is expressed on yeast in a functional format suitable for screening.*

Stability of the passaged library was validated by spiking it first with 5% cells expressing 4-4-20 scFv (which binds fluorescein.) The library was amplified 100-fold and frozen, in each of five repeated cycles, with analysis of fluorescein-binding and epitope tag expression after each cycle. Remarkably, following a total library amplification of $10^{10}$-fold, we observed almost no change in any of the expression parameters (data not shown). This indicates that library representation can be stably maintained during library passaging, largely due to strong glucose repression of scFv expression during amplification.

**Antibody library screening for model antigens:** The functional diversity of our scFv library was characterized by screening it against approximately 20 antigens, two of which have been used for characterizing phage display scFv libraries (fluorescein and hen egg lysozyme). Other antigens that we have used to screen the library are EGFR, epidermal growth factor, other EGFR ligands, P53 and phospho peptides. Magnetic bead screening (Miltenyi) was combined with flow cytometry to screen $10^{10}$ cells. In each case, on the first attempt several clones were isolated with binding affinity ranging from 1-100 nM. Such high affinity antibodies could be used directly for many applications (e.g., Western blots, ELISA, immunofluorescence). Affinity was measured by

direct titration of yeast-displayed scFvs with biotinylated antigen. Furthermore, this approach can be readily extended to the generation of scFv antibodies to any protein that can be cloned and expressed or for which representative peptides can be synthesized.

### 3.3.2 Renewable surface-biomolecular interaction assay techniques

We have extensive experience in both renewable surface techniques [Bruckner-Lea et al., 2000; 2001; 2002a; 2002b; Chandler et al., 1999; 2000; 2001] and mass spectrometry of biomolecules [Conrads et al., 2000; 2001; Lipton et al., 2002; Smith et al., 2001; 2002; ], placing us in a unique position to combine these techniques for the identification of biomolecular complexes that bind to protein-derivatized beads. We have developed and used renewable surface systems for many applications with complex (biological and/or environmental) samples, including nucleic acid isolation and analysis [Bruckner-Lea et al., 1998; 2000; 2002; Chandler et al., 1999; 2000], cell concentration [Brukner-Lea 1999; Chandler et al., 2001], cell-based luminescent assays, automated enzyme-amplified immunoassays, and radiochemical applications. PNNL has developed several renewable surface flow cells for bead capture and release [Egorov et al., 1999; Grate et al., 1999]. These flow cells are suitable for capture and release of microparticles ranging from 5-150 µm in diameter, and ranging in consistency from rigid (glass, ceramic), to soft (Sepharose, agarose). We are currently using these renewable microcolumns with on-column UV-Vis spectroscopy to directly detect DNA-protein and protein-protein interactions on the surface of derivatized Sepharose beads, utilizing a moving piston or rotating rod flow cell. The microcolumns are nominally 1 ul in total volume, with an optical path length of 1 mm. We have also built a completely automated fluidic system for DNA-based pathogen detection that includes a renewable surface microcolumn for DNA purification and concentration from environmental samples.

Protein-protein interactions can be observed in real-time with on-column absorbance detection. The detection limit of the system was estimated to be 5 ng protein bound to the beads (ca. 2 x $10^{10}$ molecules, as measured using commercial IgG antibodies), and the binding capacity of a 3 µl column typically ranges from 35 to 350 pmoles (5 to 50 µg IgG), depending on the bead size and concentration of binding sites. Protein quantities in the ug range are routinely captured and eluted. After a binding event, it is possible to use the absorbance change at 280 nm to estimate the total amount of protein bound to the microbeads. If the protein amino acid sequence is known, the absorbance change can be correlated to quantitative amounts of protein since tyrosine and tryptophan contribute to the measured absorbance change. The resulting binding curves can therefore be used to determine the binding stoichiometry. In addition to binding stoichiometry, the shape of the binding curve contains information about the binding kinetics.

The key advantages of this approach for monitoring protein binding events are that the experiments are rapid, automated and require small sample volumes. Typical sample volumes currently used are 50 µl and the perfusion rate is 0.5 µl/s, so that sample perfusion and washing can occur in less than 5 minutes. Equilibrium ~ 3 hours. These volumes and times could easily be scaled down by using smaller syringe pumps. In addition, this approach does not require fluorescent labeling and it is suitable for the detection of protein complexes containing multiple proteins. This is in contrast to surface sensitive methods for monitoring protein binding such as surface plasmon resonance devices (e.g., BiaCore), in which the measurement sensitivity falls off exponentially with film thickness. When general non-specific binding is a concern, the spectral information can be used in some cases to distinguish between molecules, especially if specific absorbing functional groups are present. The on-column measurement of absorbance as a function of wash stringency conditions can also be used to investigate the conditions required for removing non-specifically adsorbed molecules.

Mass spectrometry by LC-MS typically requires an analysis volume of 10-100 µl, containing a total protein load of about 5-10 µg per analysis. These volumes and protein quantities are readily handled and [10]optically detected using 1-5 µl renewable surface microcolumns (as described above). LC separation of enzymatically digested proteins (peptides) typically requires about 15 minutes, so the ability to process samples in several minutes is also compatible with downstream LC-MS for protein identification.

### 3.3.3 Use of microfluidic devices for sample preparation

**Protein and peptide separations:** Researchers at ORNL have performed pioneering work in the development of microfabricated devices for the manipulation and analysis of microscale samples. This work has included microchip-based separations of proteins and peptides on time scales that are typically 1/10th to 1/100th that of conventional chromatographic and electrophoretic methods [Gottschlich et al., 2001; Liu et al., 2000a; 2000b] . The resolving power and speed of these devices is demonstrated in **Fig. 3.3.2**, which shows a 15-minute, two-dimensional analysis of fluorescently labeled tryptic peptides by tandem LC/CE separation. The peptides were separated in the first dimension by micellar electrokinetic chromatography (MEKC) in a 25-cm microchip channel. During the MEKC separation, 20-msec plugs of the eluent were diverted into a 2-cm column at 2-sec intervals for rapid CE analysis. Each eluting peak from the MEKC column was therefore analyzed at multiple points by CE. The separated components were detected by laser-induced fluorescence (LIF) and the signal intensity is shown as grayscale intensity vs. migration times in the two separation modes. (More than the expected 34 peptide spots were observed due to incomplete digestion and/or variable labeling). Electrokinetic valving in the 2D microchips allows rapid sampling of the 1st dimension eluent stream with zero dead volume, and permits comprehensive analysis of all sample components. The theoretical peak capacities for orthogonal separations in these 2D microchips are in the range of 2500-4000, comparable to 2D PAGE and conventional tandem LC/CE analyses requiring several hours [Bushey & Jorgenson, 1990; Moore et al., 1996; Opiteck & Jorgenson, 1997; Opiteck et al., 1997].

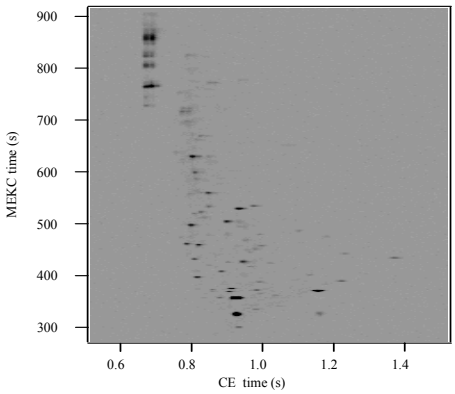
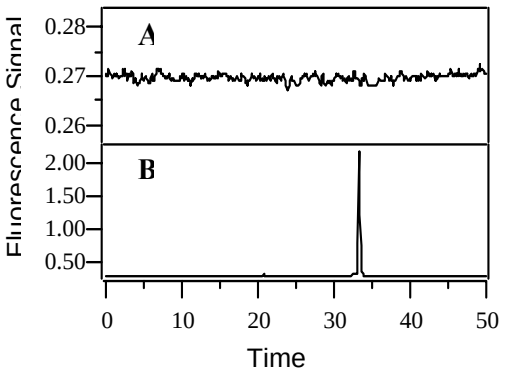



**Fig. 3.3.2.** Two-dimensional analysis of tryptic peptides by tandem LC/CE

Because MEKC and CE are not truly orthogonal separation modes, the peptides in **Fig. 3.3.3** are not distributed throughout the available 2D "space", and the effective peak capacity for this system may be only ~50% of the theoretical maximum. This would, however, be much more than adequate for the separation of tryptic peptides from most proteins and multi-protein complexes. An effective peak capacity of ~150 has been obtained by the combination of C18 reverse-phase open-channel electrochromatography (OCEC) and CE in the 2D microchip analysis of β-casein tryptic peptides [Gottschlich, 2001 #138]. The OCEC/CE combination is attractive from the standpoint of buffer compatibility with on-line mass spectrometry (see below). A drawback is the ~10-fold lower sample loading capacity for OCEC relative to MEKC. This problem could potentially be addressed by using packed reverse-phase columns in the first dimension for higher loading and peak capacities.

### 3.3.3.2 Integration of sample processing

Multiple sample processing operations may be performed on microfluidic chips and integrated with separation and detection. As examples, we have demonstrated devices that perform on-chip trypsin digestion of proteins followed by electrophoretic separation and post-column fluorescent labeling of the peptides [Gottschlich et al., 2001]; PCR amplification of DNA followed by preconcentration and size separation of the reaction products [Khandurina et al., 1999]; and isolation of single, fluorescently-labeled mammalian cells followed by electrically-induced cell lysis and separation of intracellular fluorescent metabolites (**see 3.3.3.3** below). The on-chip concentration of biomolecules is achieved by means of a microfabricated porous membrane that allows passage of buffer electrolytes but traps larger migrating molecules in an injection chamber prior to separation [Khandurina et al., 1999]. Signal enhancements of >1000 have recently been achieved by this method for the detection of fluorescently labeled proteins (**Fig. 3.3.3**).

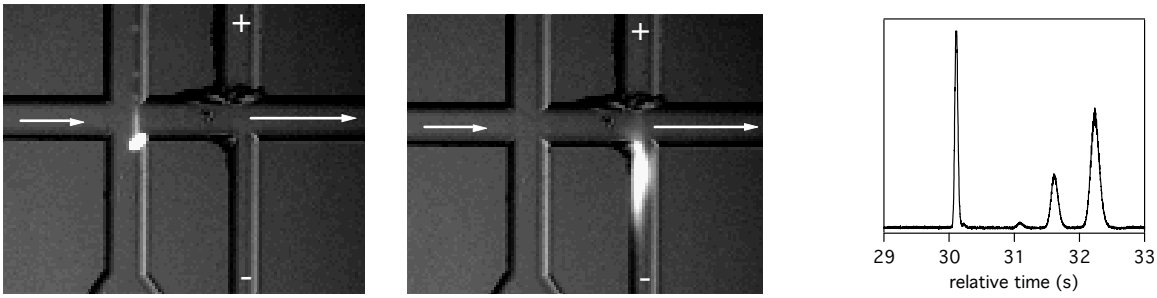


#### 3.3.3.3 Coupling to mass spectrometry

Microchips may be interfaced with mass spectrometry using electrospray ionization (ESI) [Ramsey & Ramsey, 1997]. Devices with replaceable nanospray tips have been constructed at ORNL and used for rapid MS analysis [Lazar et al., 1999; 2001]. The tip is attached perpendicularly to the plane of the microchip at the end of the separation channel and the voltage to establish the spray is applied at a fluid reservoir. The system has some similarities to conventional nanospray sources in that it operates at very low flow rates and does not require any pressure assistance for fluid delivery. These devices have been used to perform on-chip tryptic digestion coupled with ESI-time-of-flight (TOF) MS analysis of peptides for a number of proteins [Lazar et al., 2001]. Digestions were performed at ambient temperature in electrosprayable buffers containing 0.1-2 μM protein and protein/trypsin ratios of (1-20)/1 (w/w). Sequence coverage ranged from 70-97% for 10-15 min digestions. Subattomole quantities of peptides from nanomolar concentration samples have been detected with the device using TOFMS with high-speed data acquisition and storage rates [Lazar et al., 1999]. This level of sensitivity would allow detection of unique peptides in 10 μg/ml digests of protein complexes up to MW = $10^7$. Higher sensitivities might be achieved by on-chip concentration of proteins or peptides as described above.

#### 3.3.3.3 Single cell analysis

The channels in microfluidic devices are appropriately sized to efficiently handle and transport cells. In addition, the volumes injected into the separation channels on these devices are similar to the volumes of single cells. This makes microfluidic devices amenable to single cell analysis. Therefore, we have developed microfluidic devices that are capable of handling and rapidly lysing single cells and electrophoretically separating any fluorescently labeled contents. These devices consist of a double cross structur and the cells are hydrodynamically pumped through the main (horizontal) channel. All of the channels are coated with PDMS and a pluronic polymer to prevent the cells from sticking to the walls. At the first cross solubilizing agents are added to keep the hydrophobic constituents of the cell in solution when the cell is lysed. The cell is lysed at the second cross by the electric field that is applied perpendicular to the direction of cell transport. About 50 percent of the fluorescently labeled contents are injected into the separation channel. The electric field is applied as a DC biased square wave at 75 Hz and ranges from 450 to 750 V/cm. To test this device Jurkat cells have been labeled with Oregon green diacetate, a cell membrane permeable dye. The cell hydrolyzes the dye to Oregon green which is cell membrane impermeable and forms several other metabolic byproducts. These products are detected by laser induced fluorescence detection ~ 1 cm downstream of the cross. Several cells can be detected over the course of a 60 sec run giving rise to throughputs of ~ 10 cells/min. All of the cells analyzed in any single run demonstrated consistent peak patterns.




Left panels: Image of cell being transported through the channel manifold toward the syringe pump as indicated by the arrows. B. Image of cell being lysed as it comes into contact with the electric field applied perpendicular to the direction of cell flow in the main channel. Right panel: Separation of the Oregon green and Oregon green metabolites from the cytosol of a single cell.

### 3.4 Research Design and Methods

#### 3.4.1 Overall technical approach

The overall process of going from a sequenced genome to the characterization of thousands of protein complexes can be accomplished for almost any organism in a fairly linear series of steps, using the scFv approach. The first step is the availability of a fully sequenced genome, analyzed for the presence of likely open reading frames. From this basis the necessary PCR primers can be designed to clone each hypothetical ORF into an appropriate expression vector. The ideal expression vector would have the following characteristics: 1) propagation in a convenient host, such as *E. coli* or *S. cerevisiae*, as well as in the parent organism. 2) The cloning site would be flanked with appropriate linkers for both N-terminal and C-terminal labeling with distinct tags suitable for affinity purification or fluorescent imaging. 3) The vector would also be suitable for efficient *in vitro* translation. Any of these capabilities could then be used to generate substantial amounts of protein suitable for use as an affinity bait, for transduction into the experimental organism, or for use as an antigen in scFv generation. For generation of scFv antibodies, the protein would be produced by highly efficient *in vitro* translation reactions, as described below, and would be modified to contain markers suitable for affinity sorting of yeast scFv clones. Once specific scFv clones of sufficient affinity have been identified, they will be used in traditional immunoprecipitation experiments, or alternatively immobilized on renewable surfaces for microcolumn binding reactions. Protein complexes identified by reaction with scFv on microcolumns or in multiwell plates will then be used for mass spectrometric identification of proteins. The detailed procedures for accomplishing each of these steps are detailed below.

### 3.4.2 Specific Aim #1: To use high throughput selection of scFv antibodies in yeast to generate a comprehensive library of high affinity scFv antibodies suitable for isolation of protein complexes

#### 3.4.2.1 Cloning of *S. oneidiensis* ORFs for use as affinity tags

One objective of goal one is to compare the ability of different approaches to isolate protein complexes from *S. oneidensis* for subsequent identification using mass spectrometry. Accordingly, in the general cloning strategy described in the Core proposal, distinct sets of acceptor expression vectors will be utilized depending upon the approach chosen. For example, the expression of native or affinity tagged "bait" proteins in *S. oneidensis* would demand shuttle vectors capable of replication in both *E. coli* and S. *oneidensis* with and without various affinity tags to be created. Proteins encoded by the ORFs will also be expressed using *E. coli* as the host or a cell-free translation system. Again, the particular downstream use will dictate which promoter (T7, araBAD, trc, lac) system and affinity tag(s) should be present. These "bait" molecules will be created with numerous affinity tags

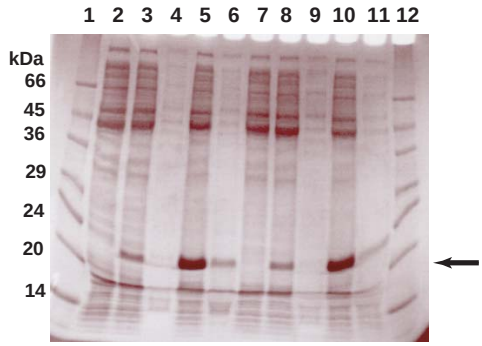



**Fig. 3.4.1.**. SDS-PAGE analysis of cell-free production recombinant chloroamphenicol acetyltransferase (CAT) utilizing RIKEN transcription/translation system. A large-scale (3 ml reaction with 30 ml feeding solution) cell-free dialysis reaction was incubated for 8 hours at 30 C and 10 µl samples were removed for SDS-PAGE (lanes 5, 10) and 30 µl for a CAT assay. Lanes 1, 12, MW markers with indicated sizes. Arrow indicates CAT protein. Lanes 2, 7 are controls without addition of CAT gene to extract. Lanes 3, 6, 8 are aliquots form small-scale reactions without feeding solutions (lane 6 was control supplied by RIKEN). Amount of total CAT estimated from gel is at least 20 mg. Enzymatic assay of CAT protein confirmed expected full activity.

facilitating tandem affinity purification of an isolated complex with reduced amount of background in relation to the same molecule with only one tag. Obviously, a number of different tags (GST, calmodulin binding peptide, FLAG, Myc, HA, V5, protein A, 6x His, chitin binding domain and maltose binding domain) and the location of the tags (N-, C-, or both termini) need to be examined to determine if any one approach will be suitable for isolation of protein complexes on a genomic scale. However, the presence of any affinity tag could prevent the formation of a complex by steric hindrance or cause the protein to misfold. Therefore, yet another approach, using single chain antibodies that specifically recognize different *S. oneidensis* proteins in order to immunoprecipitate complexes containing the target protein will be examined. In this approach, the *Shewanella* proteins will serve as antigens for the selection of the scFv antibodies, rather than as the "bait" molecules themselves. Ideally, the antigens would contain only one affinity tag (for purification) that could be removed by a protease prior to incubation with the antibody library; however, the proteins produced for the above mentioned endeavors could also be used.

In short, the sheer number of the ORFs that need to be cloned into such a multitude of expression vectors, depending on the particular downstream use, dictates using a recombination-based cloning strategy to complete this aim in a timely and cost-efficient manner.

### 3.4.2.2 Cell-free protein synthesis

Cell-free translation offers the possibility for high throughput protein production. Development of bacterial cell-free coupled transcription-translation systems to produce proteins was first described decades ago [Kim et al., 1996]. Subsequent advances exploited development of recombinant DNA technologies to produce suitable templates, engineered *E. coli* strains to minimize mRNA degradation by nucleases, efficient and readily available T7 RNA polymerase for transcription, and numerous other advances. The potential for large-sale production of recombinant proteins is currently under development by the group at RIKEN Yokohama Institute [Kigawa & Yokoyama, 1991; Kigawa et al., 1995; 1999a; 1999b; 1999c; Kudlicki et al., 1992]. As of September 2001, the RIKEN group had developed methods to attempt cell-free protein synthesis from 200 different mouse cDNAs per day. Of these 200 attempts, at least half yield 2 mg each of soluble protein. Reproduction of this extraordinary technology would enable application of the cell-free protein synthesis technology to identify and produce protein complexes. Dr. E.J. Ackerman (PNNL) learned the basic RIKEN methods during a visit to Yokohama last September, and has received DOE funding and RIKEN approval to return for an intensive visit so that the technology can be established at PNNL.

**Overview of RIKEN methods:** One of the essential components of translation extracts is the preparation of the *E. coli* S30 extracts. The *E. coli* strain is commercially available BL21 from Stratagene, grown in shaking flask cultures using standard 2XTY growth media. It is essential to harvest the cells in early log-phase, and to immediately chill the culture flasks, and then complete all subsequent processing steps at 4°C under conditions designed to minimize introduction of RNases. Breakage of the cells is a crucial step, with optimum results being obtained using a suspension of glass beads in a commercially-available Hitachi instrument. After breakage, the cells were immediately centrifuged and then dialyzed under conditions that minimize dilution of the S30 extract. The extracts can be stored in liquid nitrogen for several months.

DNA preparation using Qiagen or perhaps other methods appears to be sufficient, provided that RNAase-free procedures are followed. It is essential to produce highly concentrated T7 RNA polymerase. No suitable commercial source is available, but the appropriate clone can be obtained from the Studier lab and the enzyme is easily purified. T7 polymerase, creatine-kinase, S30 extract, amino acids, DNA template, and buffer are combined and the coupled transcription/translation reaction incubates for approximately 8 hours. An example of a translation reaction analyzed by SDS-PAGE at PNNL is shown in **Fig. 3.4.1**.

Protein purification utilizes affinity chromatography that exploits the His-Tag. For high throughput expression/purification, the RIKEN group had standardized on Beckman robots with Qiagen reagents. Without optimizing purification conditions to accommodate non-standard Qiagen buffers, at least 100 soluble mouse proteins/day are produced in quantities of 2 mg each. This represents a 50% success rate in terms of attempted

gene translations. Failures are overwhelmingly attributed to precipitated protein, rather than failed transcription or failed translation reactions.

**Additional benefits of translation extracts:**

- Purified soluble proteins can validate the single-chain antibodies.
- Mutant proteins can be expressed, including incorporation of unconventional amino acids.
- Proteins may form part of (non-antibody) bait on protein chips
- Available proteins can be used to verify suspected interactions in complexes
- Potential valuable data for identified proteins that both can and can not be expressed in soluble form; *i.e.* data base of unexpressable proteins may provide clues for why they can not be expressed.
- Available proteins will provide an important deliverable for entire program

**Intrinsic disorder in proteins:** Despite production of 100 soluble mouse proteins/day at RIKEN, this still represents 50% failure rate. One possible explanation may be that some proteins are intrinsically unstructured and remain so until binding occurs with an appropriate partner [Wright and Dyson, 1999]. There is substantial intrinsic disorder in all genomes [Dunker et al., 2000; Romero et al., 1998; 2001]. Intrinsic disorder might even be responsible for the binding diversity of proteins involved in a broad cascade of protein-protein interactions [Garner et al., 1999; Dunker et al., 2001a; 2001b]. Dunker and coworkers have developed a program to predict disorder, and its utility has been experimentally verified [Iakoucheva et al., 2001]. Analysis of *Shewanella*'s genome predicts that ~20% of the proteins contain disordered regions consisting of 40 or more consecutive amino acids. Insights provided by this analysis should help to identify proteins whose expression and/or stability may be problematic.

### 3.4.2.3 Generation of scFv libraries for *Shewenella* proteins

**Selection to generate R1 sub-library:** The scFv library is screened for protein specific affinity reagents by combining massively parallel magnetic bead enrichment technologies with high speed flow cytometric cell sorting (20,000 events/second). The starting scFv library is screened at 10 fold redundancy or coverage. This represents about 1 x $10^{10}$ yeast. This number of yeast are easily acquired from 250 mls of 2.0 $OD_{600}$ culture (2 x $10^7$cells/OD). Initial yeast cultures are done under non-inducing conditions, in order to avoid any negative selection effects of scFv expression. Prior to each sorting step, the yeast are grown overnight in media containing 2% galactose/2% rafinose/0.1% glucose to induce the surface expression of the scFv with minimal yeast growth. The yeast are pelleted, washed, and resuspended in a total volume of 10mls PBS/1% BSA/2mM EDTA.

Antigens will be produced from Shewenella ORFs expressed in *an in vitro* transcription and translation kit, which can yield up to a milligram of purified protein (as described above). A six-His tag will be used for purification of the antigen from the transcription/translation mix. 200 µg of purified antigen will be biotinylated using the EZ-Link™ Sulfo-NHS-biotinylation Kit from Pierce. The resulting purified antigen is added to yeast cultures at a

final concentration of 100 nM. For linear peptides 1uM is recommended. For a 60kd protein this represents about 60 µg of purified, biotinylated protein. 100 proteins could be screened simultaneously without putting the overall protein concentration above 1mg/ml. Antigen is added to the cells at 25$^{o}$C for 30 minutes and then put on ice for 5 minutes. The yeast are pelleted and washed twice in 50mls of 4$^{o}$C buffer before resuspension in 10mls of buffer and addition of 100 µl of anti-biotin Miltenyi Macs Beads. The yeast are incubated with the magnetic beads for 15 minutes on ice. The cells are pelleted and washed one time in 50mls of buffer. The 1-2 x $10^{10}$ yeast are put over the LS Macs column and about 5 x$10^{7}$ yeast are eluted. This represents round 1 output or the R2 sub-library.

**Sub-library expansion:** The yeast are grown in non-inducing selectable media until 1 x $10^{9}$ yeast or 100x expansion, then sub-cultured in inducing media for 12 hours at 20$^{o}$C. The yeast are prepared for a second round of selection on the Macs$^{Tm}$ magnetic column using streptavidin beads instead of anti-biotin beads and antigen at 100 nM . The eluted yeast are resuspended in buffer containing streptavidin conjugated with Alexa-488 and anti-myc monoclonal antibody for 30 minutes. Goat anti-mouse phycoerythrin mAb is used to label yeast for sorting by flow cytometry to select the top 0.1% of Alexa-488 positive yeast representing the highest affinity binding. Myc is used as an epitope tag for scFv, and is displayed on the surface of the yeast clones expressing scFv; these clones are the R3 sub-library. The R2a sub-library was sorted during selection and is no longer available. The yeast culture is expanded to 8 x $10^{7}$ yeast, representing 100-1000x coverage; this number can be easily sorted in 1 hour on a cell sorter. ScFV expression is then induced and cells are prepared for cell sorting by suspension in 100 nM antigen and anti-myc antibody. The yeast are sorted for myc positive and streptavidin positive yeast representing full length scFv clones that bind antigen. Under ideal conditions, 50% of the yeast cells surviving the selection process express scFv that bind one of the 100 antigens in the R4 sub-library. If there are obvious antigen binding clones in the R4 sub-library, that is, greater than 0.1% of total population, clone specificity can be determined at this point. The R4 sub-library can be sorted again by staining with only one of the antigens at a time. Therefore, a single antigen would be added to each of the 100 aliquots and would then be subsequently sorted. Each sort could be done in about a 5 minute time period assuming 0.1% of the cells are positive for antigen binding. We have screened for 48 antigens and sorted out antigen-binding clones for each in a 4 hour time period.

Higher affinity clones can be obtained by incubating the yeast in the presence of a lower concentration of antigen for example10 nM antigen instead of the 100 nM used in the prior round of selection. Sorting of the antigen binding clones in two directions for different levels of myc normalized antigen binding would allow multiple affinities to be sorted and identified.

**Transfer of scFv to secretion vector:** We have created a set of PCR primers that will amplify all scFv clones out of the yeast in a colony PCR and the product in combination with a prepared linearized secretion vector can immediately be transformed into yeast. After 3 days colonies can be picked and checked by colony PCR again to check if the clone contains the scFv insert.

The yeast containing the scFv clone in the secretion vector is now grown in inducing media to generate upwards of 1-10 mg of scFv antibody in volumes of 100 – 1000 ml. The scFv is secreted into the media. The yeast are pelleted by centrifugation and the supernatant is run over a 6 His column to bind the scFv. The purified scFv is eluted by adding imidazole. The purified scFv will be characterized by SDS-page for purity and concentration, and by ELISA for antigen binding and specificity of the purified scFv. Then the scFv is ready for use in affinity isolations of complexes.

.

Table 1: Idealized single chain isolation.

| Selection Round | Complexity | Frequency of epitope specific clone | Library amplification by growth | Epitope-specific clones pre-selection | % recovery -number recovered | Number of cells recovered | Enrichment |
|---|---|---|---|---|---|---|---|
| R1 full Library (beads) | $10^9$ | $1/10^9$ | 10x | $10/ 10^{10}$ | 50% - 5 | $5 \times 10^7$ | 100x |
| R2a sub-library (beads) | $5 \times 10^7$ | $1/10^7$ | 100x | $500/ 5 \times 10^9$ | 50% - 250 | $5 \times 10^7$ | 50x |
| R2b sub-library (cell sorter) | $5 \times 10^7$ | $1/ 2 \times 10^5$ | none | $250/5 \times 10^7$ | 50% - 125 | $5 \times 10^4$ | 500x |
| R3 (cell sorter) | $5 \times 10^4$ | 1/400 | 1000x | $125000/10^8$ | 50% - 62500 | $1 \times 10^5$ | 250x |
| R4 single clone analysis | $5 \times 10^4$ | 6/10 | | | | | |

### 3.4.3 Specific Aim #2: To develop a renewable surface microcolumn system compatible with isolation of protein complexes using immobilized scFv or other affinity matrixes and subsequent on-line digestion to generate peptides for mass spectrometric analysis.

The primary outcomes from this task will be new modules for automated processing of cell lysates for high throughput analysis, and a prototype system for automated processing of cell lysate solutions for mass spectrometric identification of protein complexes. Such a system will increase the experimental throughput for protein binding studies, where high throughput is needed to investigate many protein complexes and unravel the mysteries of protein function and regulation. The primary steps required for the processing of cell lysates for the mass spectrometric determination of protein complexes are shown in **Fig. 3.2.2**. Renewable surfaces are especially suited for the isolation of protein complexes (step 1) since new particles are used for each analysis to ensure that there is no sample carryover or column fouling. In addition, on-column optical detection can be used to optimize protein binding and wash conditions, determine whether samples are suitable for downstream MS analysis (enough protein load) and control downstream processing conditions. The value of renewable surfaces will also be investigated for the tryptic digestion (step 2) and solid phase extraction (step 3) modules in order to decrease carryover and improve system reliability and automation.

The general experimental approach involves using renewable surface-mass spectrometry for determining the proteins in cellular extracts that will bind to biomolecules of interest that are attached to microbeads. The experiments will involve the following steps:

- Generate protein binding sites (e.g. scFv) and attach them to microbeads.
- Pack renewable surface microcolumn with derivatized beads.
- Perfuse column with cell extracts collected from cells grown under various conditions (e.g. aerobic, anaerobic, or with alternative electron acceptor substrates).
- Monitor protein binding to beads in real-time by measuring changes in the UV-Vis absorbance spectra.
- If there is evidence for specific protein-protein interactions, elute proteins for downstream processing. Both low stringency and higher stringency washes can be analyzed.
- Tryptic digestion of proteins either on a renewable surface column, a standard column or in solution.
- Solid phase extraction either on a renewable surface column or standard column.
- Combined electrophoretic and chromatographic separation of tryptic peptides on MicroChip.
- ESI-MS/MS or FTICR identification of proteins.

#### 3.4.3.1 Selection and preparation of cell extracts and protein binding reagents

Initial testing of the operation of each separation module and integration of the modules will be conducted using known protein solutions. Experiments will then proceed with *S. oneidensis* for the capture and analysis of the biomolecular complexes. The protein complexes present will be compared for cells grown under different conditions. A detailed description of the protein complexes that will be studied are described in the CORE proposal. Cell lysates will be prepared by standard methods such as by bead beating using three 90 sec cycles at

4500 rpm in a Biospec Minibeadbeater, with a 5 min cool down on ice between cycles. Cell lysates will then immediately be placed on ice to inhibit proteolysis prior to analysis.

#### 3.4.3.2 Renewable surface-biomolecular interaction assay system

**Sequential injection-renewable surface system with on-column absorbance measuremen:** Protein binding onto derivatized microbeads will first be investigated using our current automated RS-BIA system. Briefly, this system includes a standard sequential injection system in combination with a custom PNNL rotating rod flow cell. The sequential injection system consists of a syringe pump and selection valve for automatically aspirating and dispensing microbeads, samples and reagents. The software for system control has been written in-house and is available for use. Optical fibers (400 µm diameter) are used for light delivery to the microcolumn and detection using a Spectral Instruments spectrophotometer. The volume of the microcolumn packed bed is 1 µl, and the optical path length across the column for absorbance measurement is 1 mm. An autosampler for the bead and sample delivery and collection system will be added to this set-up in order to eliminate the need for operator intervention when analyzing five or more samples in series, and multiple derivatized-bead types.

**Biomolecule and single chain antibody (scFv) attachment to microbeads:** Biomolecule binding sites (e.g., antibodies or scFv) will be attached to microbeads using established methods [Hermanson, 1996]. Two types of beads will be initially considered: spherical silica beads and Sepharose beads. We have experience using both of these bead types and they are well-suited for renewable surface protein binding experiments since they are easily handled within the fluidic RS-BIA system and have a low absorbance in the UV-Vis region that is monitored (260-800 nm). The final bead choice will depend upon the best substrate for the combined RS-BIA-MS experiments, and the substrate that results in less non-specific binding from the bacterial cell lysate. We have found that protein immobilization randomly on bead surfaces (through amine groups on proteins) can provide microcolumn binding capacities ranging from 35-350 pmoles per 3 µl column, which is suitable for this application. One primary consideration in choosing a biomolecule attachment chemistry for this application, is to choose linkers with minimal absorbance in the UV range used for detection (280 nm is used for unlabeled protein detection), since increasing the background absorbance limits the dynamic range.

The silica beads will be purchased in spherical form from Waters Corp. (YMC beads, 50 µm diameter). We have optically screened silica beads from many vendors, and found that these beads had the lowest background absorbance. The beads will be silanized to generate a surface with the desired functional group, and then reacted with the scFv polypeptide. For example, silanization with EPPTMS (3-(2,3-epoxypropoxy) propyltrimethoxysilane, Gelest Company), will result in epoxide groups on the bead surface. The epoxide can then be reacted with amine groups on a protein, resulting in random orientation of the protein on the surface. Ethanolamine can then be used to block any unreacted binding sites on the bead. Biomolecule attachment to Sepharose beads can easily be achieved by directly attaching the polypeptides to cyanogen bromide activated Sepharose 4B beads (Sigma). Ethanolamine can then be used to block the unreacted binding sites. We have successfully used both of these approaches for attaching biomolecules to beads for on-column protein binding studies [Brukner-Lea et al., 2001].

**Non-specific binding:** Both surface chemistry and solution composition are critical in order to minimize non-specific binding. Therefore, we will compare the binding onto scFv-derivatized columns to the binding onto blank columns (all reactive sites blocked with e.g., ethanolamine). We will also investigate the effect of wash stringency on the removal of protein material. Absorbance will be continually monitored during binding and wash steps, so the wash solutions can later be analyzed for protein content and also protein composition using the MS. We plan to initially analyze several wash fractions. These wash fractions will provide a binding profile ranging from non-specifically bound, to weakly bound, to tightly bound biomolecules. As we gain experience with a specific cell/protein system, we expect we will need to analyze fewer wash fractions. Our renewable microcolumns also contain built-in temperature control (from ambient to 95˚C), providing another (potential) experimental variable with which to optimize stringency and investigate protein-protein binding interactions (including protein-protein interactions in thermophilic organisms, under *in-situ* temperature conditions).

Another approach to address nonspecific binding and identify the proteins that are bound *directly* to the binding site (on the bead) is to use on-column crosslinking strategies to crosslink biomolecules to the binding site. After crosslinking, very stringent wash conditions can be used to remove all other (non-crosslinked) biomolecules. The crosslinker is then specifically cleaved, to allow release of the bound protein. This approach could be used to determine the protein attached directly to the binding site.

Our long-term aim regarding all non-specific binding studies will be to determine a systematic approach for optimizing binding and wash conditions, so that this optimization can ultimately be automated to the greatest extent possible.

**Control of downstream processing:** The renewable surface optical signal will be used for screening prior to mass spectrometry, so that only samples with positive BIA optical signals will be further analyzed by mass spectrometry for protein identification. It will be necessary to determine the range of absorbance changes during protein complex elution that provide suitable quantities of proteins for MS analysis. Initially, this will be based on absorbance changes assuming 100% efficiency of the downstream processing (and average protein content of tyrosine and tryptophan amino acids, which have significant absorbance at 280 nm), and currently desired peptide quantities for MS analysis (5-10 ng). Inefficiencies in downstream processing (such as inefficient tryptic digestion and protein adsorption to flow channels) may require set-point adjustments.

#### 3.4.3.3 System integration

This activity will involve designing, building, and testing an integrated system for automated sample preparation encompassing all of the steps from capturing the biomolecule complexes, to sample delivery to the mass spectrometer. An automated and integrated sample preparation system for the binding and analysis of protein complexes by mass spectrometry is expected to allow the automated analysis of samples at about 15 minute intervals. This will involve a significant effort in hardware, software and in developing system control algorithms (e.g., based on the on-column optical detection signal).

Engineering support for this task will be provided by the Instrument Development Laboratory (IDL) that is located in the Environmental Molecular Sciences Laboratory at PNNL, where both the Microsensors and Microfluidics, and Mass Spectrometry groups that will primarily be working on this task are located. The IDL has already built integrated fluidics systems that include renewable surface microcolumns. In addition, IDL has provided engineering support for the engineering and development of an LC-MS interface and other mass spectrometer components.

#### 3.4.3.4 Build prototype parallel processing microcolumns with optical detection

The multiplexing of multiple sample preparation systems will be important for the very high thoughput analysis of proteins using a suite of mass spectrometers. At least 2 microcolumns would be useful to allow simultaneous processing of a “control” surface to help assess non-specific binding. In addition, for some of the stages (e.g., peptide separation) it is necessary to have two systems alternately running in sample processing and cleaning modes to increase throughput. We will therefore build a 2x2 renewable surface microcolumn “array” to demonstrate the high throughput binding and elution of protein complexes for downstream analysis.

The multiplexed renewable surface system will be the size of a standard benchtop instrument. The system will include only one detector, with separate optical fibers leading to each renewable microcolumn. A multiplexer will be used to alternately measure the light output from each microcolumn (as is routinely done on microplate readers, for nearly simultaneous detection in real-time, in all microwells).

#### 3.4.3.5 Demonstrate parallel processing principle

We will demonstrate the parallel processing capability of a 2x2 renewable surface array by: 1) demonstrating parallel capture, optical detection, and elution using known protein solutions such as mixtures of antibodies and

secondary antibodies.; 2) automating the protein complex purification and detection from cellular extracts as described above for one column.

### 3.4.4 Specific Aim 3. Development of microfluidic devices for high throughput processing of protein complexes from microbial cells.

The purpose of this research is to develop monolithic microfabricated devices that allow automation and integration of standard chemical and biochemical operations used in the isolation and analysis of protein complexes from microbial cells[e.g., Rappsilber et al., 2002]. The conceptual design of such a device is shown in **Fig. 3.4.2**. Although this device appears complex, the individual block operations, such as cell lysis, capture of tagged protein complexes on affinity beads, etc., are relatively simple and can be performed separately and in succession until all operations are completed. Many of the individual operations described here have been previously demonstrated on microchips by us or others, including the manipulation of microbial cells [McClain et al., 2000]and paramagnetic beads [Fan et al., 1999]. Separations of proteins and peptides and on-chip proteolytic digestion with electrospray MS have been described in Preliminary Results.

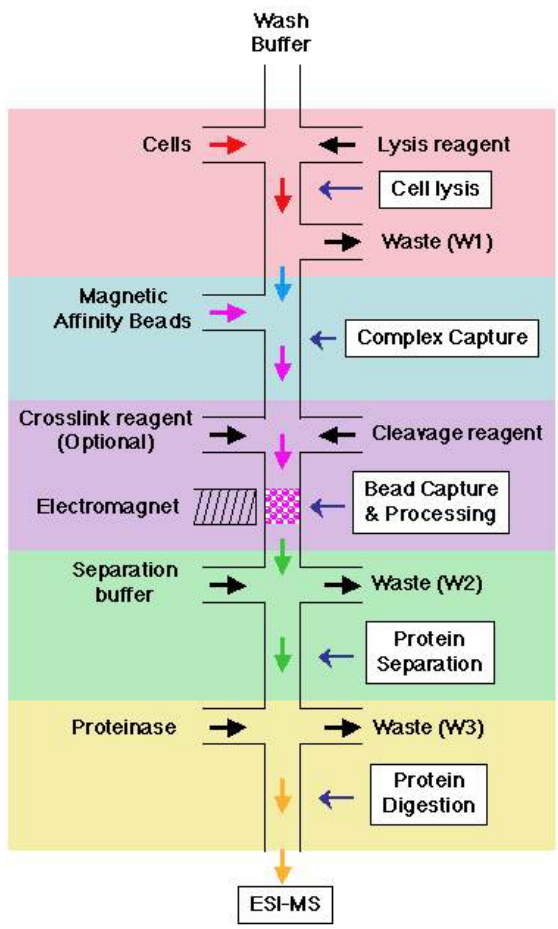



**Fig. 3.4.2.** Conceptual design of microfuidic device for integrated processing of protein complexes.

The microdevices are fabricated by standard photolithographic and wet etching techniques to create channels in planar glass substrates, and the channels are enclosed by bonding a glass coverplate to the etched surface. Fluid access to the channels is provided by holes drilled through the substrate or coverplate. Although the schematic shows channels of uniform width, channel dimensions would vary in different operational blocks to accomodate larger volumes (μL's) at the beginning of the process and reduction to smaller volumes (nL's) as processing proceeds. The channel length-to-width ratio would obviously be much greater than shown.

Fluid transport in selected block channels may be controlled by applying voltages to appropriate fluid reservoirs to induce electroosmotic flow (EOF). For example, the application of positive voltage to the reservoirs containing suspended cells and cell lysis reagents, with zero voltage (ground) applied to the waste (W1) reservoir, will cause flow of the cells and reagent toward W1 with mixing in the common channel. No electric fields are applied to the other channels during this step, so that little or no flow is generated elsewhere in the channel structure. Flow could also be generated by applying vacuum at W1, if it is desired to avoid exposing the cells or lysate to an electric field. In this case it might be necessary to fabricate the channel structure with constricted regions between operational blocks in order to restrict in-flow from other channels. Flow in the lysis channel would be stopped to allow sufficient time for complete cell lysis, and the lysate would then be mixed with magnetic affinity beads for capture of tagged protein complex. Flows for this step could be generated by applying positive voltage at the reservoirs containing wash buffer and affinity beads with ground at W2, or by applying vacuum at W1. (In the latter case, the channel layout would be altered to prevent premature mixing of beads with cleavage reagent). Beads would be collected by activation of an electromagnet placed above a specific channel region and would be held at that position while operations such as buffer washing, crosslinking (if desired) and complex release are performed by addition of appropriate reagents. Released complex proteins would then be mixed with separation buffer and injected into the separation column. The optimum separation mode will need to be determined. Although this step is conventionally performed by SDS gel electrophoresis, this mode is problematic for down-stream, on-line mass

**This page intentionally left blank.**

### 3.6 Literature Cited

Belov, M.E., G.A. Anderson, N.H. Angell, Y. Shen, N. Tolic, H.R. Udseth, and R.D. Smith, Dynamic range expansion applied to mass spectrometry based upon data-dependent selective ion ejection in capillary lc-fticr for enhanced proteome characterization. *Anal. Chem.*, **73**: p. 5052-5060 (2001).

Boder, E.T., K.S. Midelfort, and K.D. Wittrup, Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proceedings of the National Academy of Sciences* **97**(20): p. 10701-10705 (2000).

Boder, E.T. and K.D. Wittrup, Optimal screening of surface-displayed polypeptide libraries. *Biotech. Progress* **14**(1): p. 55-62 (1998).

Bruckner-Lea, C.J., E. Ackerman, B. Dockendorff, D.A. Holman, J. Kim, and J.W. Grate, Renewable surface biosensensors with optical detection. *Proceedings of the Electrochemical Society*, **2001-18**: p. 157-164 (2001).

Bruckner-Lea, C.J., D.P. Chandler, B. Dockendorff, J. Kim, and J.W. Grate, Tailoring microbead chemistry for bioanalytical separation and detection. *Anal. Chim. A.*, submitted (2002b).

Bruckner-Lea, C.J., M.S. Stottlemyre, J.W. Grate, D.A. Holman, F.J. Brockman, and D.P. Chandler, Rotating rod renewable microcolumns for automated, solid-phase DNA hybridization studies. *Anal. Chem.*, **72**(17): p. 4135-4141 (2000).

Bruckner-Lea, C.J., D.A. Holman, B.L. Schuck, F.J. Brockman, and D.P. Chandler, Strategies for automated sample preparation, nucleic acid purification, and concentration of low target number nucleic acids in environmental and food processing samples. *SPIE Proceedings*, **3544**: p. 63-71 (1999).

Bruckner-Lea, C.J., S.-S. Lee, D.A. Holman, B.L. Schuck, D.P. Chandler, and F.J. Brockman. *Automated nucleic acid isolation from soil extracts using renewable affinity microcolumns in a mesofluidic flow system*. in *53rd Northwest Regional ACS Meeting*. June 17-20, Pasco, WA (1998).

Bruckner-Lea, C.J., T. Tsukuda, B. Dockendorff, J.C. Follansbee, M.T. Kingsley, C. Ocampo, J.R. Stults, and D.P. Chandler, Renewable microcolumns for automated DNA purification and flow-through amplification: From sediment samples through polymerase chain reaction. *Anal. Chim. A.*, **in press** (2002a).

Bushey, M.M. and J.W. Jorgenson, Automated instrumentation for comprehensive two-dimensional high performance liquid chromatography/capillary zone electrophoresis. *Anal. Chem.*, **62**: p. 978 (1990).

Chandler, D.P., J. Brown, D.R. Call, S. Wunschel, J.W. Grate, D.A. Holman, L. Olson, M.S. Stottlemyre, and C.J. Bruckner-Lea, Automated immunomagnetic separation and microarray detection of *e. Coli* 0157:H7 from poultry carcass rinsate. *Int. J. Food Microbiol.*, **70**: p. 143-154 (2001).

Chandler, D.P., F.J. Brockman, D.A. Holman, J.W. Grate, and C.J. Bruckner-Lea, Renewable microcolumns for solid-phase nucleic acid separations and analysis from environmental samples. *Trends Anal. Chem.*, **19**(5): p. 314-321 (2000).

Chandler, D.P., B.L. Schuck, F.J. Brockman, and C.J. Bruckner-Lea, Automated nucleic acid isolation and purification from soil extracts using renewable affinity microcolumns in a sequential injection system. *Talanta*, **49**: p. 969-983 (1999).

Conrads, T.P., K. Alving, T.D. Veenstra, M.E. Belov, G.A. Anderson, D.J. Anderson, M.S. Lipton, L. Pasa-Tolic, H.R. udseth, W.B. Chrisler, et al., Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tages and $^{15}$n-metabolic labeling. *Anal. Chem.*, **73**: p. 2132-2139 (2001).

Conrads, T.P., G.A. Anderson, T.D. Veenstra, L. Pasa-Tolic, and R.D. Smith, Utility of accurate mass tags for proteome-wide protein identification. *Anal. Chem.*, **72**: p. 3349-3354 (2000).

Dunker, A.K. and Z. Obradovic, The protein trinity--linking function and disorder. *Nat Biotechnol*, **19**(9): p. 805-6 (2001b).

Dunker, A.K., J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, et al., Intrinsically disordered protein. *J Mol Graph Model*, **19**(1): p. 26-59 (2001a).

Dunker, A.K., Z. Obradovic, P. Romero, E.C. Garner, and C.J. Brown, Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform*, 2000. **11**: p. 161-71 (2000).

Dyson, H.J. and P.E. Wright, Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol*, **12**(1): p. 54-60 (2002).

Egorov, O.B., S.K. Fiskum, M.J. O'Hara, and J.W. Grate, Radionuclide sensors based on chemically selective scintillating microspheres: Renewable column sensor for anlysis of $^{99}$tc in water. *Anal. Chem.*, **71**: p. 5420-5429 (1999).

Eynard, N. and J. Teissie, eds. *Electrotransformation of bacteria*. Springer-Verlag: Berlin (2000).

Fan, Z.H., S. Mangru, R. Granzow, P. Heaney, W. Ho, Q.P. Dong, and R. Kumar, Dynamic DNA hybridization on a chip using paramagnetic beads. *Anal Chem*, **71**(21): p. 4851-4859 (1999).

Figeys, D., Y. Ning, and R. Aebersold, A microfabricated device for rapid protein identification by microelectrospray ion trap mass spectrometry. *Anal. Chem.*, **69**: p. 3153-3160 (1997).

Garner, E., P. Romero, A.K. Dunker, C. Brown, and Z. Obradovic, Predicting binding regions within disordered proteins. *Genome Inform Ser Workshop Genome Inform*, **10**: p. 41-50 (1999).

Gottschlich, N., C.T. Culbertson, S.C. Jacobson, and J.M. Ramsey, Integrated microchip device for the digestion, separation, and postcolumn labeling of proteins and peptides. *J. Chromatogr. B*, **745**: p. 243-249 (2000).

Gottschlich, N., S.C. Jacobson, C.T. Culbertson, and J.M. Ramsey, Two-dimensional electrochromatography/capillary electrophoresis on a microchip. *Anal. Chem.*, **73**: p. 2669-2674 (2001).

Grate, J.W., O.B. Egorov, and S.K. Fiskum, Automated extraction chromatographic separations of actinides using separation-optimized sequential injection techniques. *Analyst*, **124**: p. 1143-1150 (1999).

Grebe, T.W. and J.B. Stock, The histidine protein kinase superfamily. *Advances in Microbiology and Physiology*, **117**: p. 139-227 (1999).

Hermanson, G.T., *Bioconjugate techniques*. San Diego, CA, USA: Academic Press. 785 (1996).

Holt, L.J., C. Enever, R.M. de Wildt, and I.M. Tomlinson, The use of recombinant antibodies in proteomics. *Current Opinion in Biotech.* **11**(5): p. 445-449 (2000).

Iakoucheva, L.M., A.L. Kimzey, C.D. Masselon, J.E. Bruce, E.C. Garner, C.J. Brown, A.K. Dunker, R.D. Smith, and E.J. Ackerman, Identification of intrinsic order and disorder in the DNA repair protein xpa. *Protein Sci*, **10**(3): p. 560-71 (2001).

Kato, M., et al., Insights into multistep phosphorelay from the crystal structure of the c-terminal hpt domain of arcb. *Cell*, **88**: p. 717-723 (1997).

Khandurina, J., T.E. McKnight, S.C. Jacobson, L.C. Waters, R.S. Foote, and J.M. Ramsey, Integrated system for rapid pcr-based DNA analysis in microfluidic devices. *Anal. Chem.*, **72**(13): p. 2995-3000 (2000).

Khandurina, J., S.C. Jacobson, L.C. Waters, R.S. Foote, and J.M. Ramsey, Microfabricated porous membrane structure for sample concentration and electrophoretic analysis. *Anal. Chem.*, **71**(9): p. 1815-1819 (1999).

Kigawa, T., T. Yabuki, and S. Yokoyama, [large-scale protein preparation using the cell-free synthesis] [in process citation]. *Tanpakushitsu Kakusan Koso*, 1999. **44**(4 Suppl): p. 598-605 (1999b).

Kigawa, T., T. Yabuki, Y. Yoshida, M. Tsutsui, Y. Ito, T. Shibata, and S. Yokoyama, Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett*, **442**(1): p. 15-9 (1999c).

Kigawa, T., T. Yabuki, Y. Yoshida, M. Tsutsui, Y. Ito, T. Shibata, and S. Yokoyama, Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett*, **442**(1): p. 15-9 (1999a).

Kigawa, T., Y. Muto, and S. Yokoyama, Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for nmr analysis. *J Biomol NMR*, **6**(2): p. 129-34 (1995).

Kigawa, T. and S. Yokoyama, A continuous cell-free protein synthesis system for coupled transcription-translation. *J Biochem (Tokyo)*, **110**(2): p. 166-8 (1991).

Kim, D.M., T. Kigawa, C.Y. Choi, and S. Yokoyama, A highly efficient cell-free protein synthesis system from escherichia coli. *Eur J Biochem*, **239**(3): p. 881-6 (1996).

Kudlicki, W., G. Kramer, and B. Hardesty, High efficiency cell-free synthesis of proteins: Refinement of the coupled transcription/translation system. *Anal Biochem*, **206**(2): p. 389-93 (1992).

Kutter, J.P., S.C. Jacobson, N. Matsubara, and J.M. Ramsey, Solvent-programmed microchip open-channel electrochromatography. *Anal. Chem.*, **70**(15): p. 3291-3297 (1998).

Lazar, I.M., R.S. Ramsey, S. Sundberg, and J.M. Ramsey, Subattomole-sensitivity microchip nanoelectrospray source with time-of-flight mass spectrometry detection. *Anal. Chem.*, **71**(17): p. 3627-3631 (1999).

Lazar, I.M., R.S. Ramsey, and J.M. Ramsey, On-chip proteolytic digestion and analysis using "wrong-way-round" electrospray time-of-flight mass spectrometry. *Anal Chem*, **73**: p. 1733-1739 (2001).

Li, L., C.D. Masselon, G.A. Anderson, L. Pasa-Tolic, S.-W. Lee, Y. Shen, R. Zhao, M.S. Lipton, T.P. Conrads, N. Tolic, et al., High throughput peptide identification from protein digests using data-dependent multiplexed tandem fticr mass spectrometry coupled with capillary liquid chromatography. *Anal. Chem.*, **73**: p. 3313-3322 (2001).

Lipton, M.S., L. Pasa-Tolic, G.A. Andrerson, D.J. Anderson, D.L. Auberry, J.R. Battista, M.J. Daly, J. Fredrickson, K.K. Hixson, H. Kostandarithes, et al., Global analysis of *deinococcus radiodurans* r1 proteome using accurate mass tags. *PNAS*, Submitted (2000).

Liu, Y., R.S. Foote, S.C. Jacobson, R.S. Ramsey, and J.M. Ramsey, Electrophoretic separation of proteins on a microchip with noncovalent, postcolumn labeling. *Anal. Chem.*, **72**(19): p. 4608.4613 (2000a).

Liu, Y., R.S. Foote, C.T. Culbertson, S.C. Jacobson, R.S. Ramsey, and J.M. Ramsey, Electrophoretic separation of proteins on microchips. *J. Microcol. Sep.*, **12**(7): p. 407-411 (2000b).

McClain, M.A., C.T. Culbertson, S.C. Jacobson, and J.M. Ramsey, Flow cytometry of *escherichia coli* on microfluidic devices. *Anal. Chem.*, **73**(21): p. 5334-5338 (2000).

Mizuno, T., Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of escherichia coli. *DNA Research*, **4**: p. 161-168 (1997).

Moore Jr., A.W., J.P. Larmann Jr., A.V. Lemmo, and J.W. Jorgenson, Two-dimensional liquid chromatography-capillary electrophoresis techniques for the analysis of proteins and peptides, in Methods in enzymology, W.S. Hancock, Editor. Academic Press: New York. p. 401-419 (1996).

Oleschuk, R.D., L.L. Shultz-Lockyear, Y.B. Ning, and D.J. Harrison, Trapping of bead-based reagents within microfluidic systems: On-chip solid-phase extraction and electrochromatography. Anal Chem, **72**: p. 585 (2000).

Opiteck, G.J., K.C. Lewis, J.W. Jorgenson, and R.J. Anderegg, Comprehensive on-line lc/lc/ms of proteins. *Anal. Chem.*, **69**: p. 1518-1524 (1997).

Opiteck, G.J. and J.W. Jorgenson, Two-dimensional sec/rplc coupled to mass spectrometry for the analysis of peptides. *Anal. Chem.*, **69**: p. 2283-2291 (1997).

Ramsey, R.S. and J.M. Ramsey, Generating electrospray from microchip devices using electroosmotic pumping. *Anal. Chem.*, **69**(6): p. 1174-1178 (1997).

Rappsilber, J., S. Siniossoglou, E.C. Hurt, and M. Mann, A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal Chem*, **72**(2): p. 267-275 (2002).

Rocklin, R.D., R.S. Ramsey, and J.M. Ramsey, A microfabricated fluidic device for performing two-dimensional liquid-phase separations. *Anal. Chem.*, **72**(21): p. 5244-5249 (2000).

Romero, P., Z. Obradovic, C.R. Kissinger, J.E. Villafranca, E. Garner, S. Guilliot, and A.K. Dunker, Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomput*, p. 437-48 (1998).

Romero, P., Z. Obradovic, X. Li, E.C. Garner, C.J. Brown, and A.K. Dunker, Sequence complexity of disordered protein. *Proteins*, **42**(1): p. 38-48 (2001).

Ruzicka, J., Bioligand interaction assay by flow injection absorptiometry using a renewable biosensor system enhanced by spectral resolution. *Analyst*, **123**: p. 1617-1623 (1998).

Ruzicka, J. and L. Scampavia, From flow injection to bead injection. *Anal. Chem.*, **71**(7): p. 257A-263A (1999).

Ruzicka, J. and E.H. Hansen, Flow injection: From beaker to microfluidics. *Anal. Chem.*, **72**(5): p. 212A-217A (2000).

Shen, Y., R. Zhao, M.E. Belov, T.P. Conrads, G.A. Anderson, K. Tang, L. Pasa-Tolic, T.D. Veenstra, M.S. Lipton, and R.D. Smith, Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Anal. Chem.*, **73**: p. 1766-1775 (2001).

Siegel, R.W., B. Allen, P. Pavlik, J.D. Marks, and A. Bradbury, Mass spectral analysis of a protein complex using single-chain antibodies selected on a peptide target: Applications to functional genomics. *J. Mol. Bio.*(2): p. 285-293 (2000).

Smith, R.D., L. Pasa-Tolic, M.S. Lipton, P.K. Jensen, g.A. Anderson, Y. Shen, T.P. Conrads, H.R. Udseth, R. Harkewicz, M.E. Belov, et al., Rapid quantitative measurements of proteomes. *Electrophoresis*, **22**: p. 1652-1668 (2001).

Smith, R.D., G.A. Anderson, M.S. Lipton, L. Pasa-Tolic, Y. Shen, T.P. Conrads, T.D. Veenstra, and H.R. Udseth, An accurate mass tag strategy for quantitative and high throughput proteome measurements. *Proteomics,* in press (2002).

West, A.H. and A.M. Stock, Histidine kinases and response regulator proteins in two-component signaling systems. *TRENDS in Biochemical Sciences*, **26**: p. 369-376 (2001).

Wright, P.E. and H.J. Dyson, Intrinsically unstructured proteins: Re-assessing the protein structure- function paradigm. *J Mol Biol*, **293**(2): p. 321-31 (1999).

Xu, Q. and A.H. West, Conservation of structure and function among histidine-containing phosphotransfer (hpt) domains as revealed by the crystal structure of ypd1. *Journal of Molecular Biology*, **292**: p. 1039-1050 (1999).

**This page intentionally left blank.**

### 3.8 Milestones

The principal goal and deliverable for the initial 2-year period of this research will be the integration of protein separation and digestion steps with on-line ESI-MS. Complexes isolated by conventional or microcolumn techniques will be used in these initial studies. Peptide separations are relatively well developed and the integration of 1D and 2D peptide separations with on-line MS will be pursued as a second goal for the analysis of tryptic digests of isolated complex proteins.

| Tasks | FY03 | FY04 | FY05 |
|---|---|---|---|
| 3.4.4.1 Separation of protein complex subunits with on-line MS | x | x | |
| 3.4.4.2 Separation of subunit digests with on-line MS | x | x | |
| 3.4.4.3 Integration of post-column digestion step | x | x | |
| 3.4.4.4 Integration of protein complex capture step | | x | x |
| 3.4.4.5 Integration of cell lysis step | | | x |

**This page intentionally left blank.**