Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Mass Spectrometry 238 (2004) 119–130
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

An integrated strategy for the discovery of drug targets by the analysis of protein–protein interactions

John M. Peltier*, Srdjan Askovic, Robert R. Becklin, Cindy Lou Chepanoske, Yew-Seng J. Ho, Vladimir Kery[1], Shuping Lai[1], Tahmina Mujtaba[1], Mike Pyne[1], Paul B. Robbins[1], Moritz von Rechenberg, Bonnie Richardson, Justin Savage, Peter Sheffield, Sam Thompson, Lawrence Weir[2], Kartika Widjaja[2], Nafei Xu[3], Yuejun Zhen[4], J. Jay Boniface

Prolexys Pharmaceuticals Inc., 2150 West Dauntless Avenue, Salt Lake City, UT 84116, USA

Received 25 October 2003; accepted 8 December 2003
Available online 27 September 2004

Abstract

Proteomics-based technologies have the potential to accelerate the development of drugs, but such technologies must be well integrated in order to have a positive impact. We describe, herein, a multi-step process for the discovery of protein–protein interactions. It is shown that process stages are interdependent and can influence, either positively or negatively, subsequent steps. Optimization of each step, in the context of the full process, is essential for the overall success of the experiment.

Keywords: Protein interactions; MS/MS; Affinity tags; Drug discovery; Drug targets

1. Introduction

In the past two decades, drug companies have embraced a variety of new technologies to help accelerate the development of new drugs. These technologies have spanned the drug discovery process from the target discovery stage to late stage drug development. Examples of such technologies include: DNA sequencing, gene expression profiling, rational drug design, automated combinatorial and parallel synthesis and high throughput adsorption, distribution, metabolism and excretion (ADME) analysis. As a group, these techniques share a common underlying feature, namely, they have made it possible for drug companies to perform certain aspects of the drug discovery process at a much higher rate of throughput than was previously possible.

In the "post-genomic" era, the focus has shifted to the potential of proteomics (the systemic study of all expressed proteins in a given biological system in a given state [1]) to improve the efficiency of the drug development process. The promise of proteomics technologies is that they will have an impact on many aspects of the drug discovery pipeline. For example, proteomics has had an impact on drug target identification [2–4] and the analysis of protein interaction networks and protein expression profiles can aid in target selection and validation. A more detailed analysis of a protein structure and its interaction with small molecules, whether by computational methods or crystallography, can facilitate lead generation [5]. Furthermore, lead optimization, pre-clinical development and clinical trials can all benefit from the discovery and analysis of protein biomarkers. Such biomarkers

* Corresponding author. Tel.: +1 801 303 1792; fax: +1 801 303 1701.
E-mail address: jpeltier@prolexys.com (J.M. Peltier).
[1] PNNL, Richland Washington, WA 99352, USA.
[2] BioMicro Systems, Inc., 1290 West 2320 South, Suite D, Salt Lake City, UT 84119, USA.
[3] Myraid Genetics Inc., 320 Wakara Way, Salt Lake City, UT 84108, USA.
[4] Lilly Research Laboratories Eli Lilly and Company Indianapolis, IN 46285-1533, USA.

1387-3806/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2003.12.039

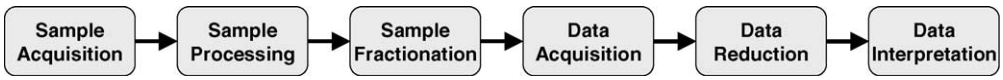

Fig. 1. Virtually any proteomic analysis is comprised of the above relatively standard process steps. Sample is acquired from a suitable biological source and preliminary processing is performed, such as homogenization or chemical modification. Then the sample is usually fractionated to focus the analysis on the components of interest. Data is acquired and reduced to remove noise or false positive signals. Ultimately, the data is interpreted in the context of the biological question that was originally asked.

might serve as early indicators of compound-based toxicity, drug efficacy or drug side effects [6]. Ultimately, a detailed understanding of the correlation between biomarkers and the side effect or efficacy profile of a drug candidate can be used to minimize clinical trial risks and develop personalized medicine [7,8].

If proteomics is to reach any significant portion of the expectations that have been ascribed to it, then the technologies in question must be deployed and integrated in such a way that they work in a seamless fashion. A high level of integration is required by the complexity of the problems being addressed and by the sheer volume of data that modern analytical tools generate. Regardless of the particular set of technologies used, without sufficient integration, valuable sample and information can be lost. Worse yet, the contamination of a sample or the misinterpretation of a result can lead to a false conclusion.

There are a wide variety of technologies, which may be grouped under the heading of proteomics. However, most techniques in proteomics have some relatively standard components that are outlined in Fig. 1. In spite of the diversity of the technologies encompassed by proteomics, currently the two most prominent approaches may arguably be classified as falling into one of two categories, namely, differential expression analysis and interaction analysis. A thorough comparison of the relative merits of these approaches is beyond the scope of this document; however, while each approach has its strengths and weaknesses they should not be considered mutually exclusive. Each has the potential ability to link specific proteins to a disease state or process. In the case of protein expression analysis, the process is discovery oriented, since there is little control over what will be found in a given analysis. Conversely, interaction analysis, as described in this paper, is a very targeted approach to finding proteins that are of relevance to a disease process or metabolic pathway. To the extent that something is already known about the pathway of interest, the later approach may be more efficient in finding proteins that are relevant to the drug discovery process.

A method that is now relatively standard for the analysis of protein interactions involves the use of an affinity-tagged bait protein, which may be used to isolate, purify, and concentrate protein complexes of interest. The isolation/purification process may be performed in a single step or in multiple steps [9–11]. Tandem purification strategies have been successfully employed to isolate protein complexes for large-scale protein interaction studies [12,13]. In this paper, we describe an optimized process for the discovery of protein–protein interactions. The basic process is outlined in Fig. 2, it is comprised of five main elements in addition to the finer details of the process. Those elements are: (1) production of the "bait protein", (2) generation of a set of prey proteins, (3) affinity isolation of the interacting proteins, (4) analysis of the isolated proteins, and (5) identification of the analyzed proteins. Step four, as employed in this study encompasses two separate approaches that are illustrated diagrammatically in Fig. 3. The system described herein has not only been optimized for maximum performance of each of the many individual steps in the process, but also to ensure that the integration of the overall process is efficient. Decisions concerning system integration included the consideration of how well particular cloning methods, purification methods, separation schemes, and mass spectrometry (MS) systems, as well as, software systems and tools fit together.

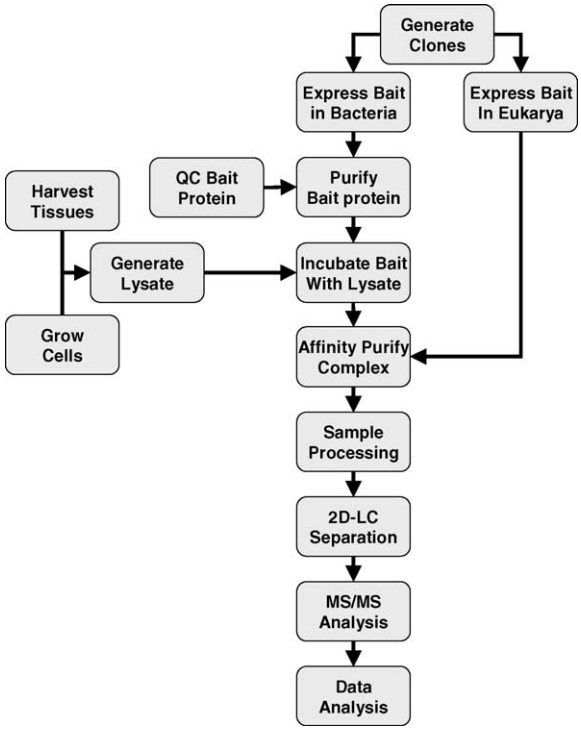

Fig. 2. An overview of the process for the analysis of protein complexes. From the generation of clones, to ultimate analysis of data, the whole process goes through the steps illustrated above. Finer details have been omitted, but each step needs to be optimized to fit efficiently with those other elements of the process with which they interact. Such interactions are not necessarily limited to the nearest neighbors in the diagram.

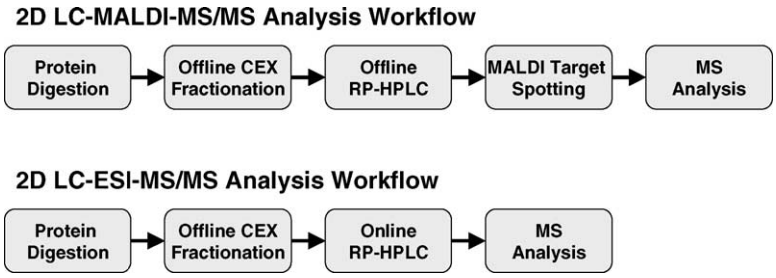

Fig. 3. A diagrammatic overview of the two sample analysis workflows employed in this study.

2. Experimental

2.1. Chemicals

Sequencing-grade tryspin was purchased from Promega (Madison, WI). All protein standards and alpha-cyano-4-hydroxy-cinnamic acid (CHCA) were obtained from Sigma–Aldrich (St. Louis, MO). CHCA was re-crystallized from water before being used and the proteins were used as provided. Acetonitrile (ACN; Optima grade) was purchased from Fisher Scientific International Inc. (Pittsburgh, PA), and trifluoroacetic acid (TFA) was purchased from J.T. Baker (Phillipsburg, NJ). The water used in this study was purified, using a Milli-Q water purification system from Millipore (Bedford, MA). HPLC grade formic acid and acetic acid were purchased from EM Science (Gibbstown, NJ). All other chemicals and reagents not otherwise listed were also purchased form Sigma–Aldrich.

The 4700 Proteomics Analyzer calibration mixture (4700 mix) was obtained from Applied Biosystems (Foster City, CA). The standard preparation of the calibration mixture contains des-Arg-Bradykinin (1.0 pmol/μl), Angiotensin I (2.0 pmol/μl), Glu-Fibrinopeptide B (1.3 pmol/μl), ACTH (1–17 clip) (2.0 pmol/μl), ACTH (18–39 clip) (1.5 pmol/μl), and ACTH (7–38 clip) (3.0 pmole/μl). A 10-fold dilution of the standard preparation was used as the calibrant for the liquid chromatography matrix-assisted laser desorption/ionization (LC–MALDI) analyses.

2.2. Cell growth and lysis for solution-based pull-downs

Large scale cultures of mammalian cells (2–25 L) were grown, in suspension, to densities of $2–5 \times 10e^6$/mL or ∼80–90% confluency for adherent cultures. Once the suspension cultures reached their desired density, the cells were centrifuged at $3000 \times g$ for 10 min. The resulting cell pellet (or plate of adherent cells) was washed 1 × with cold phosphate buffered saline (PBS). The PBS was removed and cold lysis buffer was added at a volume of 10 μL/mg of wet cell weight or 1 mL per 15 cm plate of adherent cells. The lysis buffer consists of 25 mM hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES), pH 7.5, 150 mM NaCl, 1% NP-40, 10 mM $MgCl_2$, 1 mM ethylene-diamine tetra acetate (EDTA), 10% glycerol, 1 mM dithiothreitol (DTT) and a protease inhibitor cocktail added just prior to application (1 tablet/10 mL of buffer, Roche, Mannheim, Germany). Once the lysis buffer was added, the suspension cell pellet was re-suspended with gentle pipetting and slow speed vortexing. The adherent cells were lysed (on ice) by adding the buffer directly to the culture plate and scraping the cells off of the dish. The lysates were then transferred to centrifuge tubes and allowed to incubate at 4 °C for 15 min. At 2-min intervals during the 4 °C incubation, the lysates were gently resuspended with mild shaking and/or slow speed vortexing. Following the incubation, the lysates were centrifuged at 13,000 rpm for 15 min at 4 °C to remove insoluble debris. The supernatant was then aliquoted in fresh centrifuge tubes, snap frozen in liquid nitrogen and then stored at −80 °C. Protein concentration in the resulting lysate was determined, using a Bradford colorimetric assay from Bio-Rad (Hercules, CA). Functional integrity of the lysate was checked, using one or more of the following assays: (a) a histone deacetylase (HDAC) activity assay to assess nuclear protein activity; (b) an alkaline phosphatase assay to assess cytoplasmic protein function; or (c) western blotting for expression of specific proteins of interest, such as but not limited to HDAC.

2.3. Transfection

Each plasmid was transfected into 293 T-Rex cells (Invitrogen Corporation-Carlsbad, CA), using the following method: 24 h prior to transfection, cells were plated at 90% density in 10 cm dishes. For DNA complex formation and transfection, 24 μg of the expression plasmid DNA and 60 μL of lipofectamine 2000 reagent (Invitrogen Corp.) were separately diluted in 1.5 mL of serum free media. After a 5-min incubation period at room temperature, the DNA and transfection reagent dilutions were combined, mixed gently and allowed to incubate for 20 min at room temperature. The lipofectamine/DNA mixture was then added to the media of 293 T-Rex cells at ∼90% density in a 10 cm plate, rocked gently and placed in an incubator at 37 °C (5% CO_2). Following a 4-h incubation period, the media was changed and the cells were placed back in the incubator. Twenty-four hours after the transfection, 250 μg/mL of genetecin antibiotic

was added to the culture media to eliminate non-transfected cells.

2.4. Protein expression and harvesting for cell-based pull-downs

Seven days after transfection and multiple passaging of the cell populations to permit sustained expansion, transgene expression was de-repressed by adding tetracycline at a final concentration of 10 μg/mL for a period of 48 h. Addition of tetracycline displaces the tet repressor protein (expressed by the T-Rex cells) and allows transcription from the CMV promoter. Following the period of induction, the media was removed, the culture plates were rinsed once with cold PBS and protein lysates, containing the expressed-tagged genes encoded by the expression vectors were harvested, using the gentle lysis buffer (described above) which was added directly to the plates. The lysed cells were removed from the dishes, using cell scrapers and the extracts were placed on ice for 15 min. Following this incubation, the lysates were centrifuged at 13,000 rpm for 15 min to remove cell debris and insoluble matter. The remaining supernatant was aliquoted and snap frozen in liquid nitrogen and stored at −80 °C.

2.5. Pull-down purification

One milliliter of 5 mg/mL HEK-293 whole cell lysate was incubated at 4 °C for 1.5 h rotating with affinity beads bound to 30 μg of the tagged bait protein. The samples were briefly centrifuged to pellet the beads and the lysate was removed. Beads were transferred to clean Eppendorf tubes and washed three times with 1 mL of buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 10% glycerol, 0.1% NP-40). Bound complexes were eluted in three fractions of 75, 75, and 100 μL of pull-down buffer, containing 0.2% *N*-lauroylsarcosine, and the eluates were combined. The eluates from bait, containing a HIS tag were spiked with 10 mM imidazol and incubated with a 10 μL bed of Ni-NTA agarose (Invitrogen) for 30 min at 4 °C to reduce the amount of bait in the sample. The supernatant from the nickel beads was collected and a 5 μL aliquot of each sample was analyzed by SDS–PAGE, while the remainder was ethanol precipitated for mass spectrometric analysis.

2.6. Protein complex separation for MS analysis

2.6.1. Protein digestion

The protein digestion procedure has been described elsewhere [14]. Briefly, 5–10 μg of total protein was denatured with 25 μL of 8 M urea/0.2 M NH_4HCO_3 solution. Subsequently, 5 μL of 45 mM DTT was added, and the tube was incubated at 60 °C for 15 min. After cooling to room temperature, 5 μL of 100 mM iodoacetamide was added to the tube and the solution was kept at room temperature, in the dark for 15 min. The urea was diluted to a final concentration of 2 M by the addition of 60 μL of H_2O. Trypsin was added to the tube at a protein/enzyme ratio of 20:1 by weight, and the solution was incubated at 37 °C for at least 2 h.

2.6.2. Peptide desalting

Tryptic peptides were desalted, using a peptide MicroTrap cartridge (Michrom BioResources, Auburn, CA) before being loaded onto a strong cation-exchange (CEX) column or a reverse-phase HPLC (RP-HPLC) column. The trap was conditioned with the addition of 100 μL of reagent A (0.5% acetic acid + 0.1% TFA in water) followed by 100 μL of reagent B (0.1% TFA in 95% ACN), and was then equilibrated twice with 100 μL of reagent A. The digested sample was acidified with 10% TFA and then loaded onto the trap. The flow-through volume was loaded onto the trap one more time to ensure complete binding of the peptides. The trap was subsequently washed with 50–70 μL of reagent A, and the peptides were eluted with 50–70 μL of reagent B. The desalted sample was concentrated, using a CentriVap vacuum centrifuge (Labconco, Kansas City, MO), and was re-suspended in 5.5 μL of 0.5% acetic acid/2% ACN prior to CEX injection.

2.6.3. Cation-exchange (CEX) chromatography for MALDI tandem mass spectrometry (MS/MS)

Desalted peptides were loaded onto a strong cation-exchange column (300 μm i.d. × 50 mm, SCX polymer widepore, Vydac, Hesperia, CA) equilibrated with 0.5% acetic acid/20% acetonitrile (CEX buffer A), using an autosampler. Buffer A was 0.5% acetic acid/20% ACN, and buffer B was 250 mM ammonium acetate in 0.5% acetic acid/20% ACN. The HPLC pump flow rate was 300 μL/min, which was split down to a column flow rate of 4 μL/min. The CEX eluent was collected in two fractions (flow through and the 10% buffer B as one fraction and 100% buffer B as another fraction), using a Probot micro fraction collector (Dionex, Sunnyvale, CA). The collected fractions were concentrated, using a CentriVap vacuum centrifuge.

2.6.4. Cation-exchange chromatography (CEX) for electrospray ionization (ESI)-MS/MS

Desalted peptides were loaded onto a strong cation-exchange column (300 μm i.d. × 50 mm, SCX polymer widepore, Vydac, Hesperia, CA) equilibrated with 0.5% acetic acid/20% acetonitrile (CEX buffer A), using an autosampler. The peptides were eluted, using a 50 min gradient as described below. The elution buffer used was 250 mM ammonium acetate in 0.5% acetic acid/20% acetonitrile (CEX buffer B). The gradient was held at 0% buffer B from 0 to 10 min. It was then linearly increased to 30% buffer B in 28 min. Buffer B was again linearly increased to 100% by 30 min and held constant until 35 min when it returned to 0% for column re-equilibration. The eluted peptides were collected, using a Probot microfraction collector (Dionex) at time 0–10, 10–16, 16–22, 22–28, and 28–35 min, to yield a total of five fractions. The fractions were dried, using a CentriVap vacuum centrifuge.

2.6.5. Reverse-phase high-performance liquid chromatography separation for LC–MALDI experiments

The RP-HPLC was performed, using a nano LC system from Dionex: a 75 μm × 150 mm column, a Famos autosampler, a Switchos II system and an UltiMate binary pumping module. All samples were re-suspended in 5.5 μL of 1% TFA prior to injection and were first loaded onto a 5 mm trap cartridge for in-line desalting and then were back-eluted onto the analytical column for the separation step.

For MALDI analyses, the separation employed the following conditions: solvent A was 0.1% TFA, and solvent B was 0.1% TFA in 100% ACN; the flow rate was 250 nL/min; the gradient was 0–6.5 min, 2–14% B; 6.5–41.5 min, 14–35% B; 41.5–51.5 min, 35–55% B; 51.5–52.5 min, 55–80% B; and 52.5–56.5 min, 80% B constant. For the LC–MALDI experiments, the HPLC eluent was directly mixed with MALDI matrix at a flow rate of 800 nL/min via a Micro Tee fitting (Upchurch Scientific, Oak Harbor, WA) before being deposited onto a bar-coded blank MALDI plate (Applied Biosystems), using a Probot micro fraction collector. The CHCA MALDI matrix that was mixed with the HPLC eluent was made up at a concentration of 3 mg/mL in 70% ACN with 0.1 mg/mL of ammonium acetate. Spots were deposited every 20 s and a total of 144 spots were collected in a 12 × 12 array for each HPLC run. Six external calibration spots, containing 10-fold diluted 4700 mix, were manually spotted across the plate, with two spots in the middle and four at the four corners of the LC–MALDI plate.

2.7. Mass spectrometry

2.7.1. Matrix-assisted laser desorption/ionization mass spectrometry

Samples were analyzed, using a 4700 Proteomics Analyzer MALDI-TOF/TOF (TOF/TOF; Applied Biosystems). All MS spectra were recorded in reflector mode and were derived from the sum of 750 laser shots. MALDI-MS/MS data from the TOF/TOF was acquired using the default 1 kV MS/MS install method and air as the collision gas. MS/MS data acquisition consisted of the following steps: MS spectra were recorded from each of the six calibration spots, and the default calibration parameters of the instrument and the plate model were updated for that plate; MS spectra were recorded for all 144 sample spots on that plate (each spectrum was generated by accumulating the data from 750 laser shots, using the newly updated default calibration settings); the 144 MS spectra were subsequently analyzed, using the Peak Picker software supplied with the instrument; spectral peaks that met the threshold criteria and were not on the exclusion list were included in the acquisition list for the MS/MS portion of the experiment. The threshold criteria were set as follows: mass range, 650–4000 Da; minimum cluster area, 500; minimum signal-to-noise (S/N), 10; Peak/spot, 30; maximum precursor gap, 200 ppm; maximum fraction gap, 4. A mass filter that excluded matrix cluster ions was applied. An XML file was generated which contains the list of the precursor masses selected for MS/MS and their corresponding spot numbers. The MS/MS data acquisition employed the following additional criteria: a minimum of 750 shots (6 sub-spectra accumulated from 125 laser shots each) and a maximum of 2000 shots (16 sub-spectra) were allowed for each spectrum. The accumulation of additional laser shots was halted whenever at least 10 ions with a S/N of at least 10 were present in the accumulated MS/MS spectrum in the region from *m*/*z* 400 to 90% of the precursor mass.

2.7.2. Electrospray ionization mass spectrometry (ESI-MS)

ESI-MS/MS was performed, using one of two different instruments: either an LCQ Deca XP instrument (ThermoFinnigan, San Jose, CA) or a Q TRAP™ (Applied Biosystems, Foster City, CA). A similar RP-HPLC system as was described above for the LC–MALDI set-up was used to perform the LC–ESI-MS/MS experiments. The peptide mixtures derived from the CEX fractions were separated, using a 75 μm × 150 mm column (Dionex) equilibrated in 0.1% formic acid (or 0.5% acetic acid for the LCQ) and 2% acetonitrile flowing at 250 nL/min. Solvent A was either 0.1% formic acid in water (Q TRAP) or 0.5% acetic acid in water (LCQ). Solvent B consisted of 0.1% formic acid in ACN (Q TRAP) or 0.5% acetic acid in ACN (LCQ). Peptide separations were performed, using a linear gradient from 2 to 50% solvent B, in 50 min, at a flow rate of 250 nL/min.

The LCQ was equipped with a custom built ESI source (James Hill instrument service, Boston, MA). The RP column was connected to a PicoTip Emitter (New Objective, Woburn, MA: an uncoated, 360 μm o.d., 20 μm i.d. silica needle with a 10 μm i.d. tip). The instrument was operated in the positive mode, while scanning from *m*/*z* 300–2000. The source and capillary voltages were 2200 V and 45 V, respectively, and the capillary temperature was 170 °C. Data was acquired in a data dependent mode, using one MS scan followed by three MS/MS scans of the three most abundant peaks unless they were excluded by a dynamic exclusion window of 2.5 min.

The Q TRAP is a hybrid quadrupole linear ion trap MS/MS instrument, and was equipped with the manufacturer's Flow Nanospray ion source. The RP column was connected to a PicoTip Emitter. The instrument was operated in positive ionization mode. Needle voltage was 2300 V, declustering potential was 50 V, MS collision energy (CE) was 10 V, curtain gas (ultra-high purity nitrogen; Air Liquide, Salt Lake City, UT) was set to 20 psi, and the interface heater was turned on. Instrument control was carried out by Analyst version 1.3.1 software, and data was acquired in a data-dependent mode. Each acquisition cycle consisted of an MS scan (*m*/*z* 300–1300), a higher resolution scan of up to the two most intense ions that were present in the MS scan, and up to two MS/MS scans. Maximal acquisition cycle time was 4.6 s: 0.65 s (the sum of two, 0.325 s scans) for the MS scan; 0.4 s (the sum of two, 0.20 s scans) for the higher resolution scan; and 1.7 s (the sum of three, 0.58 s scans) for each of the two MS/MS scans. Data-dependent criteria were as follows: 1–2

of the most intense ions in the MS scan were selected for MS/MS, provided that each ion exceeded 50,000 cps, had a charge-state of 2^+ to 4^+, was greater than *m/z* 300 and less than *m/z* 1200. Ions whose charge-state could not be determined by the higher resolution scan were selected for MS/MS, whereas singly charged ions were not subjected to MS/MS. Ions that were chosen for MS/MS fragmentation were subjected to a rolling collision energy that selected the CE based on the *m/z* of each ion, with a maximal CE setting of 80 V. Former target ions were excluded from MS/MS for 45 s after one repeat.

2.8. Protein identification

2.8.1. Database construction

To construct the databases used for protein identification, the following steps were performed: The NCBInr protein sequence FASTA file was downloaded, the gi numbers were updated, and the missing or incorrectly annotated taxonomies were fixed by referencing them to the NCBI taxonomy index (index of gi number versus species). The human subset of proteins in the database was extracted into a separate database called HumanNR. All of the protein sequences in the HumanNR database were matched to the corresponding protein in the latest RefSeq database, using BLAST [15]. For easy retrieval of all of the information, the corrected NCBInr file and the HumanNR file, along with their BLAST matches, were stored in an Oracle database. The Mascot server was then updated by creating a FASTA file from the relevant Oracle database entries and subsequently loading the new FASTA file onto the server, using the Mascot utilities that were provided by the manufacturer.

2.8.2. Database searching

MS/MS data obtained from the TOF/TOF, Q TRAP and LCQ were searched, using Mascot (Matrix Sciences, London, UK). All searches were performed against either the corrected NCBInr protein sequence database or the HumanNR database described above. GPS Explorer (Applied Biosystems) was used for submitting data acquired from the TOF/TOF for database searching. Data from the LCQ or Q TRAP was submitted to the Mascot server via custom software that combined peak lists from separate CEX fractions prior to submission to the Mascot server. The Mascot-based search was performed, using the default settings for the specific instrument type as supplied by Matrix Science, except that ions with scores below 10 were excluded from the results.

3. Results and discussion

3.1. Production of the bait protein

Regardless of the limitations of *Escherichia coli* as an expression system for human proteins, this bacterial system is still the workhorse of the recombinant protein world and represents a good system for rapidly and easily producing the tandem tagged proteins that are used in our laboratory. The success or failure of expression is dependent upon the sequence of the particular protein of interest (Fig. 4); protein yields will vary, as will the purity of a particular expression product. Whatever the expression level of a given protein, the presence of the tandem affinity tags on that protein can facilitate its bulk purification.

Occasionally, *E. coli* will not express a soluble version of a given protein. There are also more subtle factors than expression yield that need to be taken into consideration. Even if the protein is produced in a soluble form, it must be folded in such a way that it can be used for subsequent interaction experiments. Furthermore, given that some interactions require the presence of a specific post-translational modification (PTM), it is possible that a protein produced in a prokaryotic system may be well expressed and folded, but ineffective as a bait. In all such cases, an alternate expression system is required.

The relative ease of producing proteins, using an *E. coli*-based expression system, makes this system the first choice for most of the proteins expressed in our laboratory. To facilitate the use of alternative expression strategies, we have adopted the use of the Invitrogen Gateway™ cloning system [16]. The Gateway™ system makes possible the rapid conversion of clones into multiple destination vectors for in-

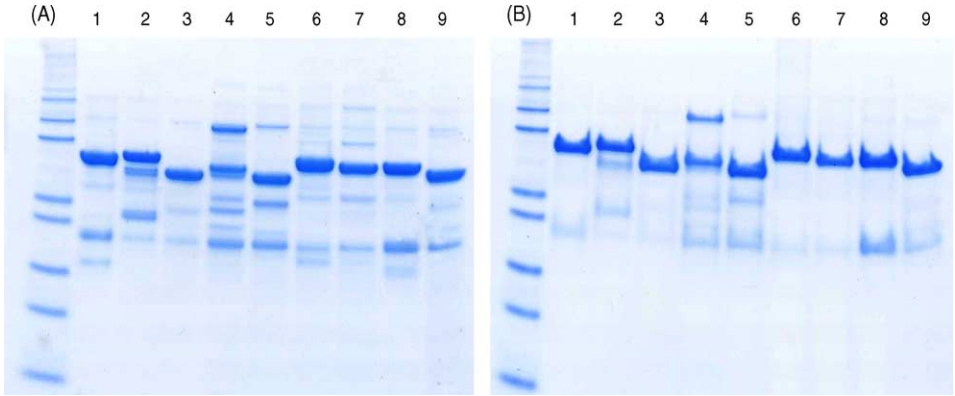

Fig. 4. Tandem-affinity purification of double-tagged baits: 5 μg aliquots of nine different baits are shown after (A) the first affinity step purification and (B) after the second affinity step purification.

troduction into a variety of expression systems, in particular for expression in mammalian and insect cell lines. This level of flexibility makes it possible to continue to use *E. coli* as the primary expression system, while having alternative eukaryotic expression systems available as a back up.

The more subtle problems represented by the second and third situations described above (folding and PTM's) have lead us to adopt a combined strategy for their resolution. If there are no obvious signs of problems with the bait protein integrity after expression in *E. coli*, as determined by a QC step, then the bait in question is moved into the pull-down pipeline. After MS analysis and data reduction, the output is analyzed to determine if there are any known interactors of the bait found in the output. Known interactors may be determined from literature reviews or orthogonal experiments, such as a yeast-2-hybrid analysis [17]. If the bait is involved in a well-characterized pathway or complex and no known interactions are identified, then the data is scrutinized to see if any confidently identified proteins are likely to participate in the same pathway as the bait protein. If none of the above conditions is met, then the open reading frame (ORF) represented by the bait in question can be moved into a vector for expression in a eukaryotic cell line. This strategy requires that the output of the pull-down analyses can be quickly correlated with a database of interacting proteins that has been curated from other sources of interaction information.

Table 1 shows data from the analysis of a pull-down in which the protein proliferating cell nuclear antigen (PCNA) was used as the bait. This protein appeared to express reasonably well (≥90% purity and >7 mg/L yield) in the *E. coli* system, yet no known interacting proteins were found in the pull-down. However, the same protein when expressed in stably transfected HEK293 cells and subjected to essentially the same pull-down process, yielded a number of known interactors.

It should be noted that there is a significant difference in the application of the pull-down process to the two expression systems. In the case of *E. coli*-derived baits, the protein is produced and then isolated on beads, and subsequently, the bead bound baits are incubated with lysate from a given cell line to allow the complexes to form. This confers two significant advantages: the initial capture of the bait proteins on beads simplifies their subsequent handling and also helps to stabilize the proteins for longer term storage. In the case of the mammalian expression system, both the bait protein and the complexes are formed endogenously, which has implications for the pull-down process that will be described below.

Table 1
A comparison of known interactors found for PCNA in solution-based vs. cell-based pull-downs

Interactors found in a cell-based pull-down	Interactors found in a solution-based pull-down	RefSeq ID	Reference
Replication factor			
C2	None	NM_002914	[21]
C3	None	NM_002915	[21]
C4	None	NM_002916	[21]
CDK2	None	NM_001798	[22]
CDC2	None	NM_033379	[23]
Ku80	None	NM_021141	[23]
PRKDC	None	NM_006904	[24]
Ku70	None	NM_001469	[23]

3.2. Creation of the lysate for pull-downs

The proteins that serve as the "prey" for a given bait protein start out "locked" in a cell or cell compartment, and must be released from the cellular or organellar membrane in order to be harvested as part of a complex which will be submitted for subsequent MS-based identification. The challenge is that the lysis conditions must be sufficiently mild that neither the secondary nor tertiary structure of the proteins is significantly disturbed; otherwise, the interactions of the members of a protein complex may never take place. In the solution-based pull-down method, where bait proteins are first produced in *E. coli* (or an alternative suitable host), a delicate balance must be achieved. This balance requires that the components of any endogenous complexes derived from the lysed cells first dissociate and then reform on the immobilized bait protein. Alternatively, free forms of the ligands may form a complex on the bait, de novo. The formation of the complexes around the "tagged" bait protein is an equilibrium process that can be facilitated by the use of a large excess of bait. In the case of the eukaryotic expression system, it is the endogenous complexes that form on the cell-based tagged bait that are the target of the analysis, and therefore, the lysis conditions must preserve these "native" complexes.

The lysates for the solution-based pull-downs are produced in bulk in advance of the experiments and then frozen in aliquots for future use. This facilitates having a ready supply of lysate for multiple pull-down experiments whenever it is required. Generally, each pull-down is done in multiple replicates and comparative experiments are performed in order to better understand the biology of the discovered interactions. Consequently, a banked source of lysate simplifies the logistics of co-ordinating pull-down experiments with lysate production. There is, however, an additional benefit to this strategy. The combination of the freeze–thaw cycle that the banked lysate aliquots go through, and the effect of the lysis buffer, can dissociate the native complexes in such a way that the complexes, along with any previously unbound interactors, can be re-capitulated on the immobilized baits. The effectiveness of this approach is illustrated in Table 2, which lists the known interactors (see references in the table) of growth factor receptor bound protein 2 (GRB2) found in a solution-based pull-down, which used tagged GRB2 as the bait.

In the case of the cell-based pull-downs, the experiment is designed to isolate endogenous complexes with minimal denaturation. Consequently, the addition of a freeze–thaw step in the process decreases the efficiency of the isolation

Table 2
Known interactors of GRB2 identified in a solution-based pull-down

Interactors found	RefSeq ID	Reference
c-cbl	NM_005188	[25]
Dynamin 2	NM_004945	[26]
WASP	NM_003941	[27]
WIRE	NM_016453	[28]
Cas-Br-M-b	NM_004351	[25]
SOS	NM_005633	[29]
DOCK180	NM_001380	[30]

of protein complexes. In fact, Table 3 illustrates that when the pull-down experiment is performed with "fresh" lysate, the number of known interactors (see references in the table) found for PCNA and *N*-ethylmaleimide-sensitive factor attachment protein alpha (NAPA) increases, relative to the case where the lysate was frozen before the pull-down was performed. Consequently, the pull-down procedure must be optimized to be both rapid and gentle to accommodate the cell-based pull-down strategy.

3.3. Affinity isolation of interacting proteins

It may be argued that the solution-based pull-down system is quite unnatural compared to the normal cellular environment in terms of the relative concentrations of the bait and interactors, as well as, the fact that the interacting proteins are no longer in their normal cellular compartments. However, a distinct advantage of the solution-based pull-down approach is that the bait can be maintained in a large in excess, which may help bias the binding kinetics in favor of the capture of the lower-affinity and lower-abundance interactors of the bait protein. While effective at capturing binding partners that may not otherwise be observed, the high abundance of bait in the solution-based pull-downs has a consequence for the down stream analysis. The pull-down elution conditions generally result in the elution of a significant amount of the bait protein. Unless steps are taken to reduce the amount of bait in the final eluted product, the presence of the bait can affect the ability of the MS systems to detect low abundance

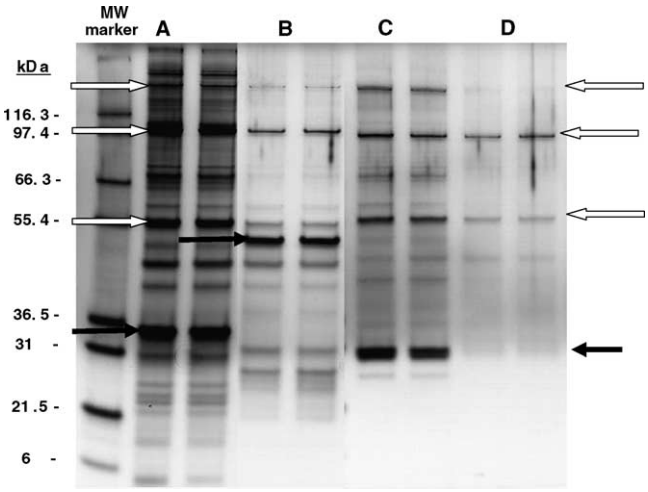

Fig. 5. An SDS–PAGE gel showing a comparison of interactors, cleanliness, and the amount of bait from duplicate Grb2 pull-downs under different processing conditions and/or tags. The black arrows show bait bands. Differences in molecular weight of the bait bands are from differences in the tag and/or cleavage. The white arrows follow the appearance and intensity of interacting proteins. (A) Single affinity purification. (B) Single affinity purification with a different tag. (C) Protease cleavage of the complex from beads with bait reduction. (D) tandem-affinity purification. (*Note*: the gel image was compiled from multiple lanes of a single larger gel from which additional lanes have been removed for the sake of clarity).

components in the sample. A strategy for dealing with excess bait and other unwanted proteins in the elution is described below.

Regardless of whether a pull-down is performed, using the solution-based or cell-based process, an important aspect of the experiment is that the beads efficiently retain the bait and its binding partners without exhibiting a significant amount of non-specific binding. Any non-specifically bound proteins that are subsequently eluted with the protein complex may complicate the interpretation of the data. Consequently, both the nature of the tags and the elution conditions affect the quality and purity of the eluted pull-down sample. Fig. 5 illustrates the effect of elution conditions and the use of various tags on the complexity of the eluted sample. In lane A and B, the difference in the complexity of the eluted material is

Table 3
A Comparison of known interactors found in cell-based pull-downs with fresh vs. frozen lysate

Bait protein	Interactors found	Fresh lysate	Frozen lysate	RefSeq ID	Reference
NAPA	NAPG	X	X	NP_003817.1	[31]
	NSF	X		NP_006169.1	[32]
	v-SNARE	X		NP_006361.1	[32]
PCNA	Replication Factor C2	X	X	NM_002914	[21]
	Replication Factor C3	X	X	NM_002915	[21]
	Replication Factor C4	X	X	NM_002916	[21]
	CDK2	X		NM_001798	[22]
	CDC2	X		NM_033379	[23]
	Ku80	X	X	NM_021141	[23]
	PRKDC	X	X	NM_006904	[24]
	Ku70	X	X	NM_001469	[23]

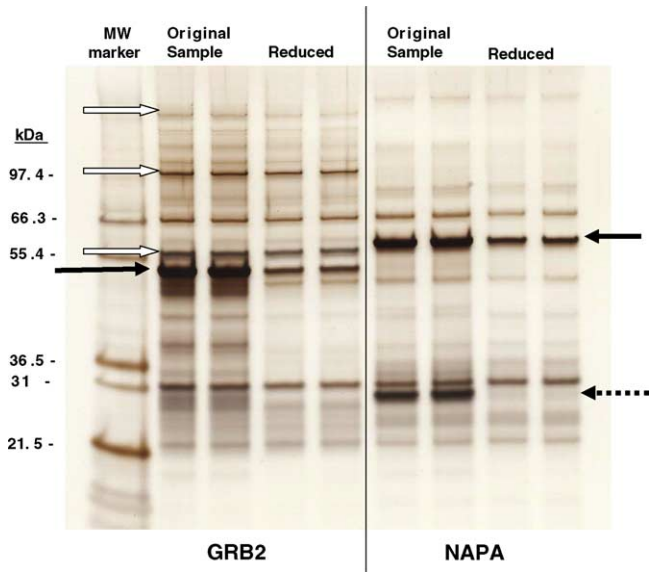

Fig. 6. Bait reduction in pull-down samples: SDS–PAGE gel showing GRB2 and NAPA pull-downs before and after bait reduction. Solid arrows indicate a bait band; the dashed arrow indicates an N-terminal fragment of a bait. Excision and analysis of individual gel bands confirmed that the bait was reduced without a significant loss of interacting proteins (white arrows on GRB2 pull-down).

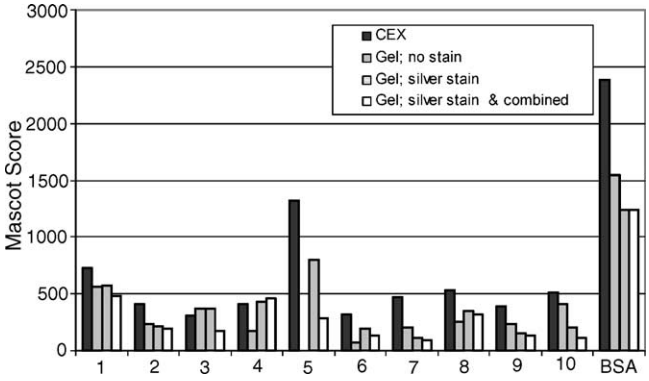

Fig. 7. A comparison of a 2D LC-gel strategy vs. a 2D LC strategy for protein analysis. A 2 μg sample, containing a mixture of proteins purchased from Sigma (55% BSA and 45% of nine other proteins: 1, alcohol dehydrogenase; 2, α-lactalbumin; 3, carbonic anhydrase; 4, cytochrome c; 5, glycogen phosphorylase; 6, hemoglobin α-chain; 7, hemoglobin β-chain; 8, lysozyme; 9, myoglobin; 10, ovalbumin) was analyzed by different 2D strategies. The samples were either processed in first dimension by cation-exchange chromatography (CEX) or various 1D gel methods with subsequent RP LC–MS/MS analysis. The 1D gel methods have been described [33]. The gel was either processed without staining (Gel, no stain); processed by extracting silver stained bands corresponding to individual proteins (gel; silver stain); or processed by combining the individual stained bands into four fractions for analysis (gel; silver stain and combined).

a consequence of a difference in the tag. The amount of non-specifcally bound proteins and bait protein is substantially reduced, while the specific interactors (as illustrated by the white arrows) are retained. To further enhance the purity of the isolated complex, a two-step purification strategy may be used. This so-called tandem-affinity purification (TAP) strategy employs the use of two orthogonal affinity tags, one of which may be cleaved from the bait after the first affinity step, by means of a specific protease. The TAP process has been successfully used in a number of large-scale experiments [12,13]. Lanes C and D show further enhancements in purity resulting from the TAP method. An adaptation of this process and its effects are also illustrated in Fig. 6. In our laboratory, the two-step purification process has been adapted to remove excess bait, as well as, improve the purity of the isolated complexes. The gel image in Fig. 6 shows that after reduction of the total amount of bait and other “impurities” the eluted sample is much cleaner without significant loss of specific interactors. This “bait reduction” step can significantly improve MS identification of the pull-down components, since the MS/MS system will spend less time analyzing peptides from the bait and consequently, is more likely to acquire data from low abundance components of the sample.

3.4. Analysis of the isolated proteins

The large-scale analysis of protein interaction networks, using a pull-down-based strategy has been previously described [12,13]. In both of those studies, a two-dimensional (2D) sample separation strategy was employed that relied on polyacrylamide gel electrophoresis for the first dimension of separation. The use of gels for protein separation is a process that is prone to contamination by keratin and often results in poor recovery of low abundance proteins or very high or low molecular weight proteins. Furthermore, the processing of samples, using gels, is labor-intensive and time-consuming. A 2D-LC strategy can, in principle, address many of the aforementioned concerns. Fig. 7 shows a comparison, performed in our laboratory, of a gel-based approach versus a 2D-LC strategy. The sample in question was constructed with a high abundance of BSA (55% of the total protein by mass) to mimic the presence of a high abundance of bait in a pull-down. In a rigorous set of comparative experiments, which employed multiple replicates, and three separate gel-based protocols; none of the gel-based analyses yielded the same level of confidence in protein identification as the 2D-LC approach.

In spite of the advantages of a 2D-LC strategy, there are other considerations that will affect the number of proteins identified. If a 2D-LC strategy is coupled only to an MS instrument that uses ESI as the mode of ionization, then a certain subset of the peptides in the sample will not be ionized or detected well. Conversely, a MALDI-based analysis will yield a somewhat different set of protein identifications. Consequently, combining the use of both ESI and MALDI for the analysis of pull-downs will yield a more comprehensive set of identified proteins than would be possible with either technique alone. The advantages of combining ESI-based analysis with a MALDI-based approach in protein identification have been described by a number

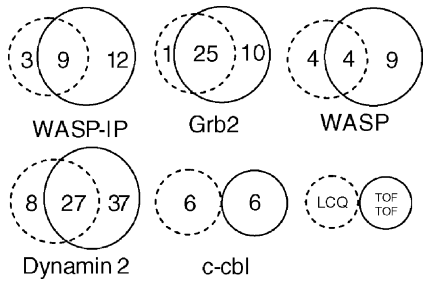

Fig. 8. A comparison of the overlap in peptides identified, using LC–ESI/MS/MS and LC–MALDI-MS/MS for GRB2 and several of its known interactors. Improved coverage is achieved when the data from each type of analysis is combined.

of laboratories [14,18]. The improved coverage afforded by this combined strategy is illustrated in Fig. 8, which shows the overlap in coverage for a number of known GRB2 interactors. In this case, a GRB2 pull-down was analyzed by both 2D LC–MALDI-MS/MS and 2D LC–ESI-MS/MS, and the combined approach is clearly more comprehensive than either approach alone. An additional benefit of this combined strategy is that proteins that were detected in both approaches are identified with increased confidence.

3.5. Identification of the analyzed proteins

The identification of proteins, as described in this paper, is wholly dependent on the comparison of MS/MS data with a relevant protein sequence database in order to accurately identify the proteins found in an analyzed sample. This process is now relatively common [19,20]; however, an often-overlooked component of the process is the reference database itself. The ideal database is one that is comprehensive, yet non-redundant. If the database is not comprehensive, then a given peptide or protein may not be identified because there is no entry in the database for an MS/MS spectrum to be matched against. Alternatively, if every possible record of protein sequence, including slight variations, is entered into the database, then multiple hits can be returned for every protein, which in turn complicates the data analysis. The problems in question are illustrated below.

Fig. 9 illustrates the consequence of not having a sufficiently comprehensive database. In this case, the sequence of human Cyclin B1 is compared to the mouse ortholog. In spite of having high sequence identity (78%, BLAST expectation value 0.0), very few tryptic peptides are shared in common by these proteins. If short tryptic peptides that contain no unambiguous sequence information are ignored (i.e., peptides

```
>gi|90411|pir||JH0509   cyclin B - mouse  Length = 430

 Score =  665 bits (1716), Expect = 0.0
 Identities = 340/432 (78%), Positives = 365/432 (84%), Gaps = 3/432 (0%)

Query: 1   MALRVTRNSKINAENKAKINMAGAKRVPTAPAATSKPGLRPRTALGDIGNKVSEQLQAKM 60
           MALRVTRN+KINAENKAK++MAGAKRVP    A SKPGLRPRTALGDIGNKVSE+LQA +
Sbjct: 1   MALRVTRNTKINAENKAKVSMAGAKRVPVTVTAASKPGLRPRTALGDIGNKVSEELQATV 60

Query: 61  PMKKEAKPSATGKVIDKKLPKPLEKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXI 120
           P+K+EAK   TGK   K LPKP+EK                                 I
Sbjct: 61  PLKREAKTLGTGKGTVKALPKPVEKVPVCEPEVELAEPEPEPELEHVREEKLSPEP---I 117

Query: 121 LVDTASPSPMETSGCAPAEEDLCQAFSDVILAVNDVDAEDGADPNLCSEYVKDIYAYLRQ 180
           LVD  SPSPMET GCAPAEE  CQAFSDVILAV+DVDA+ GADPNLCSEYVKDIYAYLRQ
Sbjct: 118 LVDNPSPSPMETCGCAPAEEYPCQAFSDVILAVSDVDADSGADPNLCSEYVKDIYAYLRQ 177

Query: 181 LEEEQAVRPKYLLGREVTGNMRAILIDWLVQVQMKFRLLQETMYMTVSIIDRFMQNNCVP 240
           LEEEQ+VRPKYL GREVTGNMRAILIDWL+QVQMKFRLLQETMYMTVSIIDRFMQN+CVP
Sbjct: 178 LEEEQSVRPKYLQGREVTGNMRAILIDWLIQVQMKFRLLQETMYMTVSIIDRFMQNSCVP 237

Query: 241 KKMLQLVGVTAMFIASKYEEMYPPEIGDFAFVTDNTYTKHQIRQMEMKILRALNFGLGRP 300
           KKM+QLVGVTAMFIASKYE+MYPPEIGDFAFVT+NTYTKHQIRQMEMKILR LNF LGRP
Sbjct: 238 KKMIQLVGVTAMFIASKYEDMYPPEIGDFAFVTNNTYTKHQIRQMEMKILRVLNFSLGRP 297

Query: 301 LPLHFLRRASKIGEVDVEQHTLAKYLMELTMLDYDMVHFPPSQIAAGAFCLALKILDNGE 360
           LPLHFLRRASK+GEVDV QHTLAKYLMEL+MLDYDMVHF PS+  +GAFCLAL+ILDNGE
Sbjct: 298 LPLHFLRRASKVGEVDVRQHTLAKYLMELSMLDYDMVHFAPSRAFSGAFCLALEILDNGE 357

Query: 361 WTPTLQHYLSYTEESLLPVMQHLAKNVVMVNQGLTKHMTVKNKYATSKHAKISTLPQLNS 420
           WTPTLQHYLSY+E+SLLPVMQHLAKNVVMVN GLTKHMTVKNKYA SKHAKISTL QLN
Sbjct: 358 WTPTLQHYLSYSEDSLLPVMQHLAKNVVMVNCGLTKHMTVKNKYAASKHAKISTLAQLNC 417

Query: 421 ALVQDLAKAVAK 432
             VQ+L+KAV K
Sbjct: 418 THVQNLSKAVTK 429
```

Fig. 9. A comparison of the sequence overlap of human and mouse Cyclin B. Alternating tryptic peptides are indicated in gray and white (e.g., XXKXXRXXK). Tryptic peptides that are common between the two sequences are indicated with a double underline.

Table 4
An example of redundant protein identification from the HumanNR database

Proteins found	gi number	Mascot score	Percent coverage (%)	Best RefSeq ID match
KIAA0820 protein	gi20521666	256	10	NP_004936.1
Dynamin 3 (Dynamin testicular) (T-dynamin)	gi27805466	256	11	NP_004936.1
DYN2_HUMAN	gi4567175	255	16	NP_004936.1
Unnamed protein product	gi21757772	255	14	NP_004936.1

with <5 amino acids), there are only six tryptic peptides in common. At the lower limits of detection of this protein in an MS experiment, only one or two peptides may be detected. If even a single amino acid in any of these peptides is incorrect in the database or mutated in the sample, then the peptide may not be identified. This is because an amino acid change can shift the peptide mass, which is an important parameter in MS-based protein identification. Furthermore, if the change involves the conversion of a lysine or arginine (the cleavage site for trypsin) to another residue, then the changes in the population of tryptic peptides derived from a protein become even more pronounced. With the natural variation in the sequence of proteins across the human population, the errors in sequencing and alterations resulting from degradation of the protein during sample handling, it may be a fallacy to say that there can be a single sequence for a given protein or that a database can be truly non-redundant. For that reason, we have chosen to use the NCBInr protein sequence database, rather than a more non-redundant database, such as RefSeq. This approach reduces the frequency of false negative identification problems, however, data analysis can be further complicated by the fact that essentially the same protein is often represented in the database multiple times under different names. For example, Table 4 illustrates the case where a single protein is represented four times. To facilitate the identification of proteins that fall into this category, we have constructed a database in which all proteins are BLASTed against the RefSeq database. Each result derived from this database (e.g., HumanNR) is annotated with the corresponding best RefSeq match and the BLAST match score; the resulting output is also illustrated in Table 4 in the column labeled Best RefSeq ID Match. While not a perfect solution, it has been our experience that this approach significantly reduces the amount of time spent determining if two database search results are, in fact, equivalent.

4. Conclusion

Proteomics technologies have the capability to be of significant benefit in efforts to accelerate the drug discovery process and reduce drug development cost. However, in order for these technologies to have a maximal impact, they must form part of a well-integrated process. For the case of a system designed to probe protein interactions, this paper illustrates many of the aspects that need to be taken into consideration for the process to work well. These aspects involve the interfaces between the different parts of the process. This includes optimization of cloning methods and expression systems for maximum flexibility in protein expression, sample processing and separations that achieve the maximum yields and purity of the samples, and analysis and database searching strategies that are suitable for the problem being studied.

References

[1] M.R. Wilkins, J.C. Sanchez, A.A. Gooley, R.D. Appel, I. Humphrey-Smith, D.F. Hochstrasser, K.L. Williams, Biotechnol. Genet. Eng. Rev. 13 (1995) 19.
[2] S. Patterson, Bioinformatics 18 (Suppl 2) (2002) S181.
[3] S. Mullner, T. Neumann, F. Lottspeich, Arzneimittelforschung 48 (1998) 93.
[4] M.A. Lindsay, Nat. Rev. Drug Discov. 2 (2003) 831.
[5] D.L. Kirkpatrick, S. Watson, S. Ulhaq, Comb. Chem. High Throughput Screen 2 (1999) 211.
[6] Y. Fujita, Nippon Yakurigaku Zasshi 116 (2000) 149.
[7] K.K. Jain, Curr. Opin. Mol. Ther. 4 (2002) 548.
[8] M. Waters, G. Boorman, P. Bushel, M. Cunningham, R. Irwin, A. Merrick, K. Olden, R. Paules, J. Selkirk, S. Stasiewicz, B. Weis, B. Van Houten, N. Walker, R. Tennant, EHP Toxicogenom. 111 (2003) 15.
[9] S. Honey, B.L. Schneider, D.M. Schieltz, J.R. Yates, B. Futcher, Nucleic Acids Res. 29 (2001) E24.
[10] O. Puig, F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, B. Seraphin, Methods (2001) 24.
[11] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, B. Seraphin, Nat. Biotechnol. 17 (1999) 1030.
[12] A.C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, A.M. Michon, C.M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.A. Heurtier, R.R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, G. Superti-Furga, Nature 415 (2002) 141.
[13] Y. Ho, A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sorensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W. Hogue, D. Figeys, M. Tyers, Nature 415 (2002) 180.
[14] Y. Zhen, N. Xu, B. Richardson, R. Becklin, J.R. Savage, K. Blake, J.M. Peltier, J. Am. Soc. Mass Spectrom. 15 (6) (2004) 803.
[15] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, J. Mol. Biol. 215 (1990) 403.
[16] J.L. Hartley, G.F. Temple, M.A. Brasch, Genome Res. 10 (2000) 1788.
[17] C.T. Chien, P.L. Bartel, R. Sternglanz, S. Fields, Proc. Natl. Acad. Sci. U.S.A. 88 (1991) 9578.

[18] W.M. Bodnar, R.K. Blackburn, J.M. Krise, M.A. Moseley, J. Am. Soc. Mass Spectrom. 14 (2003) 971.
[19] A.L. McCormack, D.M. Schieltz, B. Goode, S. Yang, G. Barnes, D. Drubin, J.R. Yates III, Anal. Chem. 69 (1997) 767.
[20] C.C. Wu, M.J. MacCoss, Curr. Opin. Mol. Ther. 4 (2002) 242.
[21] N.Y. Yao, L. Coyell, D. Zhang, R. Georgescu, J. Finkelstein, M.M. Coman, M.M. Hingorani, M. O'Donnell, J. Biol. Chem. (2003).
[22] S. Koundrioukoff, Z.O. Jonsson, S. Hasan, R.N. de Jong, P.C. van der Vliet, M.O. Hottiger, U. Hubscher, J. Biol. Chem. 275 (2000) 22882.
[23] A.S. Balajee, C.R. Geard, Nucleic Acids Res. 29 (2001) 1341.
[24] S. Ohta, Y. Shiomi, K. Sugimoto, C. Obuse, T. Tsurimoto, J. Biol. Chem. 277 (2002) 40362.
[25] S.A. Ettenberg, M.M. Keane, M.M. Nau, M. Frankel, L.M. Wang, J.H. Pierce, S. Lipkowitz, Oncogene 18 (1999) 1855.
[26] E. Lee, P. De Camilli, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 161.
[27] E. O'Sullivan, C. Kinnon, P. Brickell, Int. J. Biochem. Cell Biol. 31 (1999) 383.
[28] P. Aspenstrom, Exp. Cell Res. 279 (2002) 21.
[29] L. Sastry, W. Lin, W.T. Wong, P.P. Di Fiore, C.A. Scoppa, C.R. King, Oncogene 11 (1995) 1107.
[30] M. Matsuda, S. Ota, R. Tanimura, H. Nakamura, K. Matuoka, T. Takenawa, K. Nagashima, T. Kurata, J. Biol. Chem. 271 (1996) 14468.
[31] T.H. Brondijk, D. Fiegen, D.J. Richardson, J.A. Cole, Mol. Microbiol. 44 (2002) 245.
[32] J. Hoyle, J.P. Phelan, N. Bermingham, E.M. Fisher, Mamm. Genome. 7 (1996) 850.
[33] C. Chepanoske, B. Richardson, N. Xu, J. Savage, J.M. Peltier, Proceedings of the 51st ASMS Conference on Mass Spectrometry and Allied Topics, 2003.