We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Exploiting Protein Interaction Networks to Unravel Complex Biological Questions

Bernd Sokolowski[1] and Sandra Orchard[2]
[1]University of South Florida, College of Medicine,
[2]European Bioinformatics Institute, Wellcome Trust Genome Campus
[1]USA
[2]UK

1. Introduction

The past thirty years have witnessed a renaissance in biology as advances in technology contributed to discoveries at ever-greater orders of magnitude. One of the primary reasons for this revolution has been the advancement of technologies that allow high-throughput discovery and processing of data. This accomplishment has placed volumes of data in the realm of "discovery" science. An important point in this period came with the complete sequencing of several microbial genomes followed by the sequencing of the first multicellular organism, *Caenorhabditis elegans*, and eventually that of humans and various model organisms, such as *Drosophila melanogaster*. The edifice of the genetic code fell by wedding a biological technique developed by Sanger, known as shotgun sequencing (Sanger et al., 1977), with that of computational techniques utilizing high-speed computers. Without the advances in computer chips and processors, at a pace defined by Moore's law (Moore, 1965), sequencing would have been dramatically slower and would not have brought about the age of bioinformatics, a symbiosis of biological data, large amounts of information, and computer science.

The hypothesis that gene number is related to organism complexity is quickly discarded when comparing *Homo sapiens*, which have a genome of only 3.1 billion base pairs (Olivier et al., 2001; Venter et al., 20010), to other organisms. Estimates for the marbled lungfish, *Protopterus aethiopicus*, suggest 133 billion base pairs (Pedersen, 1971), making it the largest vertebrate genome, while, to date, the lowly amoeba, *Amoeba dubia*, is estimated to have the largest genome overall at 670 billion base pairs (McGrath & Katz, 2004). However, large genomes may be a liability, as suggested in the plant world, where *Japonica paris*, which has a genome of approximately 150 billion base pairs (Pellicer et al., 2010), grows more slowly and is more sensitive to changes in the environment (Vinogradov, 2003). In vertebrates, there appears to be an inverse correlation between genome size and brain size (Andrews & Gregory, 2009), thus, complexity may lie with other factors such as epigenetics and protein interactions. While estimates of human gene numbers rest between 20,000 – 30,000 genes, these genes may encode over 500,000 proteins. Thus, the proteome of a cell can range from several thousand proteins in prokaryotes to over 10,000 in eukaryotes. These numbers are

made more daunting with the realization that approximately 80% of the proteins in a cell do not stand alone, but rather form complexes with other proteins. Moreover, this complexity increases as the proteome of a cell can change under various conditions such as stress, disease, cell cycle, etc.

Unraveling the biological complexities in a cell's fate, growth, function, death, and disease has led to a number of techniques to unlock the mechanisms to these processes. Among these advances are those that produce large amounts of data and include gene and protein arrays, phage displays, yeast two-hybrid screens, and coaffinity or coimmunoprecipitation in combination with shotgun proteomics, which utilizes mass spectrometry. Innovations in techniques exploiting mass spectrometry are of particular significance, as this technology has improved increasingly over time. With these improvements, many technologies, once outside the realm of anyone but experts, are now user-friendly, opening the possibilities for utilization by many more scientists. Consequently, these changes in technology and accessibility have led to the formation of large databanks curated by individuals with an expertise in bioinformatics. A global view of one's own data relative to those published by others increases dramatically, as one begins to delve into these databases with software tools that retrieve large amounts of deposited data.

A number of methods can be used to unravel protein pathways, but the starting point is always the wet bench experiment that will reveal the complexity of gene expression or protein-protein interactions. This first step is fraught with potential limitations and pitfalls that vary depending on the technique. However, the object is to use an approach that will allow for the capture of many protein-protein interactions without the inclusion of too many artifacts. In my lab, we have used two techniques over the last 10 years. These are the yeast two-hybrid system and coimmunoprecipitation in conjunction with two-dimensional gel electrophoresis and mass spectrometry. Both of these procedures can yield large amounts of viable data, which leaves one with the option of either cherry-picking specific protein partners or opening a whole new world, by examining the data from a global perspective using bioinformatics. The latter approach, also known as systems biology, examines the newly discovered proteins in the larger context of protein networks or interactomes. These discoveries are made by using large datasets readily available online with software that will map the many interactions once mining of these libraries is completed and integrated with the experimental data. Within these maps you will find single proteins, or nodes, connected to other proteins, via lines known as edges. In other instances, you may find clusters of proteins in which, for example, a central protein acts like the hub in a wheel, forming connections to six or more other proteins. These hubs might reveal previously unknown functions of your protein, since they can have important regulatory roles in the cell (Fox et al., 2011). Moreover, they might suggest new subcellular functions, if they are localized primarily to specific cellular organelles.

In light of these challenges, this chapter will describe the use of bench experiments and computational techniques to determine and exploit protein-protein interactions and unravel their relation to protein networks and possibly newly discovered mechanisms. These descriptions will include coimmunoprecipitation, verification using reciprocal coimmunoprecipitation and RNAi, and data mining of specific databases. Their purpose is to provide a primer without going into specific details, since many have been described previously in great detail (e.g., Golemis & Adams, 2005; Harvey & Sokolowski, 2009; Kathiresan et al., 2009; Navaratnam, 2009).

2. Experimental design and use of bait proteins

The advances in molecular biology and protein chemistry have brought a myriad of techniques to the forefront to study molecule-molecule interactions as investigators seek to wed the relationship of their molecule of interest to various mechanisms, cycles and diseases. Among the techniques that have evolved for such studies are yeast two-hybrid screening and coimmunoprecipitation/coaffinity assays. The use of one or the other depends on which technique will reveal the biologically relevant answer and which might supply the most data for obtaining large numbers of proteins to map and build networks or interactomes. The yeast two-hybrid system does not need expensive hardware, such as a mass spectrometer, it can be done in a small laboratory, while providing high-throughput capability, and it can provide reasonably quick insights into potential binding sites. However, the system is used *in vitro* with cDNA, so any search is only as good as the quality of the screened cDNA, plus any validation of findings will occur *in vivo*, eventually. Coimmunoprecipitation combined with two-dimensional electrophoresis and mass spectrometry: allows you to pull the proteins directly from the source, since you are not dependent on obtaining a cDNA library; provides insights into protein complexes and post-translational modifications; provides amino acid sequences for potentially unknown proteins. However, further studies of interacting binding sites may need the yeast two-hybrid assay or other systems *in vitro*. Both systems generate false positives and negatives.

Yeast two-hybrid screening packages the protein of interest as cDNA in an engineered viral vector or plasmid, which is used to go fishing for other proteins that are all initially dressed as cDNA and in their own plasmid. The former is the bait, whereas the latter, known as a prey, consists of a known protein or a library of unknown proteins that are, again, in the form of cDNA (cDNA library), encoding fragments of protein derived from a tissue or organism of interest. In fusion with the bait or prey cDNAs are gene sequences that respectively encode a eukaryotic binding and activating domain. Both bait and prey can be mixed together in one soup, containing yeast cells that are transformed by the plasmids, so that many will now contain a bait and a prey cDNA. The cDNAs are incorporated into the cellular machinery and expressed as protein, after which the cells are plated on an agar-based medium. If an interaction occurs between the bait and prey proteins, the activating and binding domains interact to form a transcription factor that initiates a reporter gene, thereby changing the chromatic phenotype of the yeast for visualization. The prey cDNA, encoding a fragment of protein, is isolated from the yeast and sequenced to identify the protein involved in the interaction. Typically, this procedure involves many culture plates since the more plates the more likely you will capture a number of different interacting proteins of interest. One advantage of the yeast-two hybrid system is that you can get a fairly quick picture of the domains of interaction between the bait and prey proteins, since one of the fragments pulled from the interactions likely will be a binding site. If you do not begin with this technique for high-throughput analyses, you can use it on a low scale for studying site-directed mutagenesis in a relatively quick and reliable fashion. The downside for the yeast two-hybrid approach is that you will have to obtain a cDNA library from your tissue of interest and insert the fragments into the proper plasmid. This first step can be a weakness, because the screening is only as good as the library and proteins that are weakly or indirectly associated with your protein may be lost.
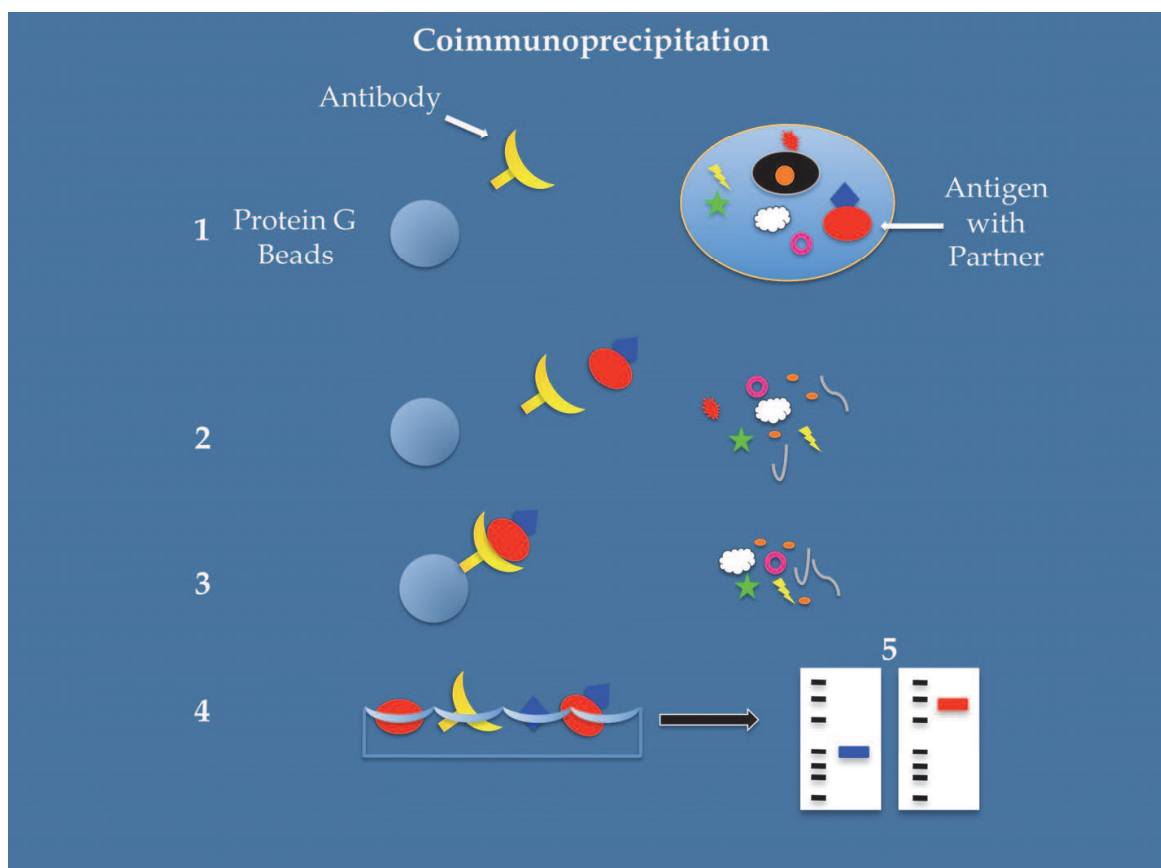
Fig. 1. Coimmunoprecipitation uses (1) a substrate consisting of protein A- or G-coated beads that bind the Fc fragment of a known antibody targeting a known antigen. (2) Once cells are homogenized the released protein lysate is mixed either with antibody alone or with antibody attached to protein-coated beads. (3) The antigen serves as bait as it brings many protein partners (prey). (4) This immunocomplex is eluted from the beads and prepared for western blotting. In high-throughput experiments, western blotting is skipped and the protein partners are separated on a 2-D gel and prepared for mass spectrometry (*see* Fig. 2). (5) For western blotting, an antibody is used to probe the blot for a known coimmunoprecipitated prey (blue). The blot can also be probed for precipitated antigen, since the known bait interacts with itself (red). In a reciprocal coimmunoprecipitation, the reverse experiment is performed, because the prey will now be used to precipitate the bait.

Coimmunoprecipitation (**Figure 1**) involves the use of an antibody on a substrate, with the antibody directed towards an antigen (the bait protein) that brings along the prey, that is, interacting proteins and protein complexes. The antibody also can be directed toward the epitope of a protein tag in fusion with a bait protein. The advantage of the technique is that you can fish in a protein soup made from your tissue or organism of interest and you can vary the antibody/antigen bait for fishing. In addition, you can pull down both direct, indirect, and weak interactions as these are highly relevant to building protein networks. The difficulty is in getting rid of interaction artifacts, so the more artifact filtering the more likely the interactions will be real. To support this effort, one can rely on a combination of centrifugation of different cellular components and using 2-D gels to better separate protein partners from one another. The components in the gel are then identified using MALDI TOF-TOF and LC-MS/MS.

3. Coimmunoprecipitation

3.1 Protein tags

The first task in setting up an experiment is to determine whether the use of cells obtained from conditions *in vivo* or *in vitro* are more suitable to the biological question at hand. At first glance, this issue may not seem relevant; however, cells obtained from conditions *in vitro* are easily accessible, allowing more freedom in the design of bait proteins and in the use of tags for quantification. A major part in this decision is determining which approach is feasible and will answer the question in a biologically relevant manner. If a system *in vitro* is chosen, there are a number of techniques that can be used, whereas the approach is more limited if cells are obtained from whole organisms or tissue lysates. For experiments *in vitro*, various heterologous expression systems are readily available, such as Chinese Hamster Ovary (CHO) and Human Embryonic Kidney (HEK) 293 cells. The increase in accessibility allows the use of isotope labeling of amino acids with Stable Isotope Labeling by Amino Acids (SILAC), Isotope-Coded Affinity Tags (ICAT), and Isobaric Tags for Relative and Absolute Quantitation (iTRAQ). These tags are successfully used in proteomic experiments involving protein-protein interactions through the differential labeling of peptides and are described extensively in a recent review (Vetter et al., 2009). Here, however, we will focus on different types of nucleotides encoding a protein tag for bait cDNA.

The cDNA of bait proteins transfected into a cell system can be epitope tagged, using FLAG (DYKDDDDK), c-myc (EQKLISEEDL), hemagglutinin (HA; YPYDVPDYA), histidine 6 (his6; HHHHHH), vesicular stomatitis virus glycoprotein (VSV-G; YTDIEMNRLGK), simian virus 5 (V5; GKPIPNPLLGLDST), and herpes simplex virus (HSV; QPELAPEDPED) tags, among others (*for additional tags see* Terpe, 2003). The FLAG tag is a hydrophobic octapeptide (Hopp et al., 1988), recognized by different anti-FLAG monoclonal antibodies (M1, M2, and M5), each with different binding and recognition characteristics. Typically this tag is used at either the N- or C-terminal ends, as is the viral hemagglutinin coat protein or HA tag. However, both can be used as an epitope tag within the C- and N-terminal domains, since tagging at the very end of either terminal may interfere with a protein-protein interaction (Duzhyy et al., 2005). Moreover, if the protein is a signaling protein, a tag at the N-terminus will be cleaved-off the main body of the protein and thus, not resolvable on a gel. These cleavage sites can be less than 20 amino acids from the N-terminus. HA tags are usually attached in multiples of two or three in fusion with a bait protein, allowing for a better signal during western blotting. The c-myc tag (Evan et al., 1985) is especially popular since there are over 150 antibodies available from different species for this particular label. In comparison, the advantage of using poly-His tags is that His binds to a chelating resin charged with metal ions such as Ni^{2+}, Cu^{2+}, or Zn^{2+} (Noronha et al., 1999; Mateo et al., 2001). It can be used to not only purify proteins, but also to bind the prey in a protein lysate poured over a bait-bound matrix in an affinity column. Once bound, the matrix-His tag can be disrupted and the prey eluted. In this scenario, lysates are used from whole organisms or tissues dissected from the organism.

3.2 Antibodies and tissue preparation

The technique for capturing protein partners is to coimmunoprecipitate protein-protein interactions using a bait antigen bound to an antibody. A second technique is to use the metal ion binding His tag in fusion with a bait protein, as mentioned above. Here, we will focus on the antibody approach, where a major hurdle is the antibody itself. These

complexes can vary not only in relation to the epitope (specificity) that is targeted, but also in relation to the affinity, which can differ by source and/or fluctuate by lot number. The first rule of thumb is that that not all antibodies are created equal. Before purchasing an antibody, check that the targeted sequence of the epitope in your protein is not similar to the sequence in a different antigen. While you might assume that this comparison was made previously, particularly if the antibody is commercially available, a quick check never hurts, as sequence databases are updated on a continual basis. Gene depositories are found at the US National Institutes of Health at http://blast.ncbi.nlm.nih.gov/Blast.cgi, the European Molecular Biology Laboratory Nucleotide Sequence Database in the UK at http://www.ebi.ac.uk/Tools/sss/psiblast/, or the DNA Data Bank of Japan at http://blast.ddbj.nig.ac.jp/top-e.html. However, all three form a consortium of the International Nucleotide Sequence Database Collaboration, so information is exchanged on a daily basis. When checking, be sure to contrast species differences; however, while these differences are not fatal, the epitope should consist of 5 – 8 amino acids that are available for binding following cell/tissue denaturation. Once these sequences are checked, initial tests using western blots are valuable to determine if the antibody recognizes the denatured target.

Prior to running a coimmunoprecipitation, a necessary step is to test the chosen precipitating antibody, because many commercial antibodies are not tested for this use. Here, the second rule of thumb is that if the antibody cannot immunoprecipitate its targeted antigen (bait), it will be useless in coimmunoprecipitating any partners (prey). Thus, checking the antibody entails doing an immunoprecipitation. The procedure is similar to a coimmunoprecipitation, but rather than probing the western blot for antigen partners, you probe for the immunoprecipitated antigen. Once verified, the antibody is suitable for use in a coimmunoprecipitation. Additionally, immunoprecipitations are useful in other applications, for example, as a control with which to compare the coimmunoprecipitated species. In this scenario, you must be working with already identified proteins. As an example, use a newly discovered partner (prey), from your high-throughput experiment to coimmunoprecipitate the bait, while also immunoprecipitating the bait as control. Both co- and immunoprecipitated species should have the same weight. This step is also referred to as a reciprocal coimmunoprecipitation (discussed in section 3.2), since it validates the original bait/prey interaction. Finally, a third use for immunoprecipitation is to increase the quantity of the antigen for western blotting and Enhanced Chemiluminescent (ECL) visualization. This technique is especially useful in pulling down lowly expressed proteins. These techniques are useful for validation following the initial high-throughput experiments.

3.3 Lysate preparation and preclearing

Once an antibody is chosen and tested, preparation of the cells/tissues for coimmunoprecipitation can begin **(Figure 1)**. A step-by-step procedure was presented previously (Harvey and Sokolowski, 2009), so here, we will just touch on the salient points and limitations. The initial preparation of the tissues for coimmunoprecipitation is critical as the quality of the protein lysate is important. The goal is to disrupt the tissue sufficiently without disrupting protein-protein interactions. Thus, lysis buffers contain anywhere from 120 – 1000 mM NaCl (less to more disruptive) as well as detergents to release hydrophobic-hydrophylic interactions. Among the reagents that can disrupt protein-protein interactions

are ionic detergents, such as sodium Deoxycholate (DOC) and Sodium Dodecyl Sulfate (SDS). However, nonionic detergents, such as Triton X-100, Tween20, Octyl β-D-Glucoside, N-dodecyl-β-D-Maltoside, Brij, Cymal, Digitonin, and NP-40, are useful in maintaining interactions. Octyl β-D-glucoside is especially helpful for releasing protein partners from lipid rafts, whereas n-dodecyl-β-D-maltoside isolates hydrophobic membrane proteins and preserves their activity. The isolation and separation of membrane proteins on a 2-D gel can be especially challenging. For example, our own initial studies, to cleanly separate BK channel partners from the membrane fraction on a 2-D gel, revealed amidosulfobetaine-14 (ASB-14), a zwitterionic detergent, as the best candidate relative to CHAPS (zwitterionic), octyl β-glucoside, and n-dodecyl β-D-maltopyranoside.

Once the tissue is dissected on ice and placed in a cold buffer with the proper protease and phosphatase inhibitors, any physical disruption is accomplished with pre-cooled equipment, on ice, and for short durations. These tissue perturbations include: mechanical disruption by grinding with a blade; liquid homogenization, by squeezing through a narrow space, as with a French press or Dounce; sonication, by using a vibrating probe to produce bubbles that burst and cause a sonic wave; or freeze/thaw, which bursts membranes via ice crystals. For minute tissues, such as the cochlea, use a 3 mm size probe to disrupt cells for 30 sec three times with one-minute intervals for cool down. Also, a simple mortar and pestle can be used and obtained in many different sizes. However, there is an art to the process, since you will not want to over-sonicate/homogenize. Such errors are reflected in mass spectroscopy results, where cytoplasmic proteins appear in the membrane fraction and vice versa. Again, as a reminder, the tube containing the tissues is kept on ice and any ensuing centrifugation should be done in either a refrigerator or a cold room.

Lysis buffers can be relatively standardized or they can vary from lab to lab with everyone swearing that theirs works the best. RIPA and Tris-HCl are commonly used lysis buffers and their ingredients can be easily found on the web with other types of buffers at sites such as http://www.abcam.com/index.html?pageconfig=resource&rid=11379#A1. However, some buffers contain metal chelating agents such as EGTA or EDTA. These chelators have the ability to bind or sequester metal ions, keeping them in solution and decreasing their activity. For example, EGTA sequesters Ca^{2+} and Mg^{2+}, but has a higher affinity for Ca^{2+} than Mg^{2+} ions, whereas EDTA binds Fe^{3+}, Ca^{2+}, Pb^{2+}, Co^{3+}, Mn^{2+}, and Mg^{2+}. The choice as to whether you add these chelators can depend on whether the protein-protein interactions you are interested in are metal ion dependent. The real differences come into play when deciding on which protease or phosphatase inhibitors to use (**Table 1**). Concentrations of these inhibitors can vary and may depend on, for example, whether or not you are interested in examining phosphorylated proteins. Protease/phosphatase inhibitors should always be mixed on the day of the experiment, since their stability varies quite a bit. Pepstatin A at a working solution of 1 μg/mL is stable for about one day, whereas the stock solution (100 μg/mL) is stable for several months. Leupeptin at 1-2 μg/mL is stable for a few hours, whereas the stock solution is stable for up to six months. Aprotinin, on the other hand is stable for about a week at 4ºC in a solution of pH 7 at a concentration of approximately 0.5-2 μg/mL. Moreover, microcystin-LR may be preferred in place of okadaic acid, as an inhibitor of protein phosphatases PP1 and PP2A, since it is more potent. The downside of using this inhibitor is that in the U.S., microcystin is on the government list of monitored reagents and, also, it is quite expensive.

Inhibitor (Synonym)	Target
AEBSF (PEFABLOC®) ($C_8H_{10}FNO_2S$)	Serine protease
ALLN (Calpain Inhibitor I, Ac-Leu-Leu-Nle-H) ($C_{20}H_{37}N_3O_4$)	Calpain 1, cysteine proteases, cathepsin B & L
Bestatin (Ubenimex) ($C_{12}H_{24}N_2O_4$)	Aminopeptidase B, leucine amino-peptidase, triaminopeptidase
Benzamidine (Benzenecarboximidamide) ($C_7H_8N_2$)	Serine proteases, trypsin-like enzymes
β-glycerophosphate (β-glycerol phosphate) ($C_3H_9Na_2O_6P$)	Serine/Threonine phosphatases
Cathepsin inhibitor I (phenylalanyl-glycyl-NHO-Bz) ($C_{26}H_{25}N_3O_6$)	Cathepsin B, cysteine protease
DFP (Diisopropyl fluorophosphate, Isoflurophate) ($C_6H_{14}FO_3P$)	Serine proteases
E-64 (trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane) ($C_{15}H_{27}N_5O_5$)	Cysteine protease, cathepsin B, cathepsin L
Na$_2$ **EDTA** (Ethylenediaminetetraacetic acid, Disodium edetate) ($C_{10}H_{14}N_2Na_2O_8$)	Chelator of Ca^{2+}, Mg^{2+}, Fe^{3+}, and other metal ions
Na$_2$ **EGTA** (Ethylene Glycol Tetraacetic Acid) ($C_{14}H_{22}N_2Na_2O_{10}$)	Chelator more specific for Ca^{2+}
Elastatinal ($C_{21}H_{36}N_8O_7$)	Elastase
EST (L-trans-epoxysuccinyl-leucylamide (3-methyl)butane-ethyl ester, E-64d, Aloxistatin) ($C_{17}H_{30}N_2O_5$)	Calpain, cysteine protease
GGACK (Glu-Gly-Arg-chloromethylketone) ($C_{14}H_{25}ClN_6O_5$)	Urokinase
Iodoacetamide (Monoiodoacetamide) (C_2H_4INO)	Cysteine peptidase
Leupeptin (N-Acetyl-L-leucyl-L-leucyl-L-argininal) ($C_{20}H_{38}N_6O_4$)	Cysteine protease
Microcystin (Fast-death factor) ($C_{49}H_{74}N_{10}O_{12}$)	Protein phosphatases type 1 (PP1) and 2A (PP2A)
NaF (Sodium Fluoride) (NaF)	Serine/Threonine and acidic phosphatases
Na$_2$ **Molybdate** (Molybdic acid, disodium salt) (Na_2MoO_4)	Phosphatases
Okadaic acid (Ocadaic acid, OKA) ($C_{44}H_{68}O_{13}$)	Protein phosphatases type 1 (PP1) and 2A (PP2A)
Na$_3$ **Orthovanadate** (Sodium pervanadate) (Na_3O_4V)	Tyrosine phosphatases
o-**Phenanthroline** (1,10-phenanthroline) ($C_{12}H_8N_2$)	Zn-dependent metalloproteinases
PAO (Phenylarsine oxide, Oxophenylarsine) (C_6H_5AsO)	Protein-tyrosine-phosphatase
Pepstatin A (Pepstatine) ($C_{34}H_{63}N_5O_9$)	Aspartic protease endothiapepsin
Phosphoramidon ($C_{23}H_{34}N_3O_{10}P$)	Membrane metallo-endopeptidase and endothelin-converting enzyme
PMSF (phenyl methyl sulfonyl fluoride; Benzenemethanesulfonyl fluoride) ($C_7H_7FO_2S$)Na$_4$	Serine proteases, fatty acid synthetase
Tetrasodium Pyrophosphate (Phosphotex, TSPP) ($Na_4O_7P_2$)	Serine/Threonine phosphatases
TLCK (Tosyllysyl chloromethyl ketone) ($C_{14}H_{21}ClN_2O_3S$)	Serine proteases
TPCK (N-tosyl-L-phenylalanine chloromethyl ketone) ($C_{17}H_{18}ClNO_3S$)	Serine endopeptidases

Table 1. Protease and phosphatase inhibitors that can be used in a cocktail mixed with a lysis buffer for protein extraction.

In order to obtain the cleanest and best protein separation on your 2-D gel, a useful step is to separate the lysate into different cellular components via centrifugation and prior to preclearing. This step is practical, especially for high-throughput experiments involving mass spectrometry (Kathiresan et al., 2009; Harvey & Sokolowski, 2009). After clearing debris, nuclei, etc., separate the membrane fraction from other soluble proteins using ultracentrifugation, by spinning the sample at 100k x g for about an hour at 4ºC. The pellet will contain membrane from the plasmalemma, mitochondrion, and endoplasmic reticulum, while the supernatant will contain any remaining soluble proteins. To obtain additional separation of organelles and various other cellular components, a necessary step is density gradient centrifugation (Huber et al., 2003). However, the initial separation of membrane and cytosolic components is useful for obtaining proteins that have undergone phosphorylation or any other changes resulting from a cell's response to cycle, developmental stage, drug response, environment, disease, etc.

Prior to preclearing and coimmunoprecipitation, a choice is made with regard to the type of beads to be used as the substrate for binding the antibody. These substrates include Protein A- or G-coated agarose, sepharose or magnetic beads. Proteins A and G bind immunoglobulins in the Fc regions of an antibody, thereby, leaving the Fab region free for antigen binding. Protein A, originally derived from *Staphylococcus aureus*, binds immunoglobulins from a number of species and has a strong affinity for mouse IgG2a, 2b, 3, and rabbit IgG. Protein G was originally derived from Group G streptococcus and tends to have an affinity for a greater number of immunoglobulins across a broader range of species and subclasses of IgG. Its affinity is strong for polyclonals made from cow, horse, sheep, and mouse IgG1. Also, Protein G has less affinity for albumin, thereby decreasing background and providing cleaner preparations. Protein A and G binding affinities for various species can be found at http://www.millipore.com/immunodetection /id3/affinitypurification. The question of agarose/sepharose or magnetic beads is a matter of choice, since arguments can be made for either one. Magnetic beads are smaller at 1 – 4 μm and provide more surface area per volume, fewer handling steps, faster protocol time, greater sample recovery, and less risk of bead inclusion in the sample. However, you need a magnetic separator. In the long run, there is likely not that much difference and the outcome will lie in performing the necessary pilot experiments.

Once the tissue is cleared of debris and nuclei, separated into different cellular components, and a choice of beads is made, begin the preclearing step. Preclearing with beads involves reducing the proportion of proteins that may bind non-specifically to the agarose/sepharose beads that are used in the coimmunoprecipitation. For high-throughput experiments, where western blots are not used, it is essential. However, if the endgame is a western blot and ECL, preclear if the background masks your protein species. One limitation of preclearing is that you may lose signal, which is especially disadvantageous if the expression of your protein is low. However, signal loss can be traced by saving non-bound components during the procedure. For preclearing, the lysate is mixed with a small volume of coated beads so that any contaminating elements that increase background noise are allowed to bind over time, usually over 30 min at 4ºC. The resultant complex of "sticky" proteins and beads are discarded (or saved for testing signal loss) after centrifugation and the supernatant is processed for coimmunoprecipitation. Preclearing is not to be confused with a bead control. Here, the cleared lysate is mixed with beads in the absence of antibody to form a non-immunocomplex, which is then processed and fractionated on a gel. An additional
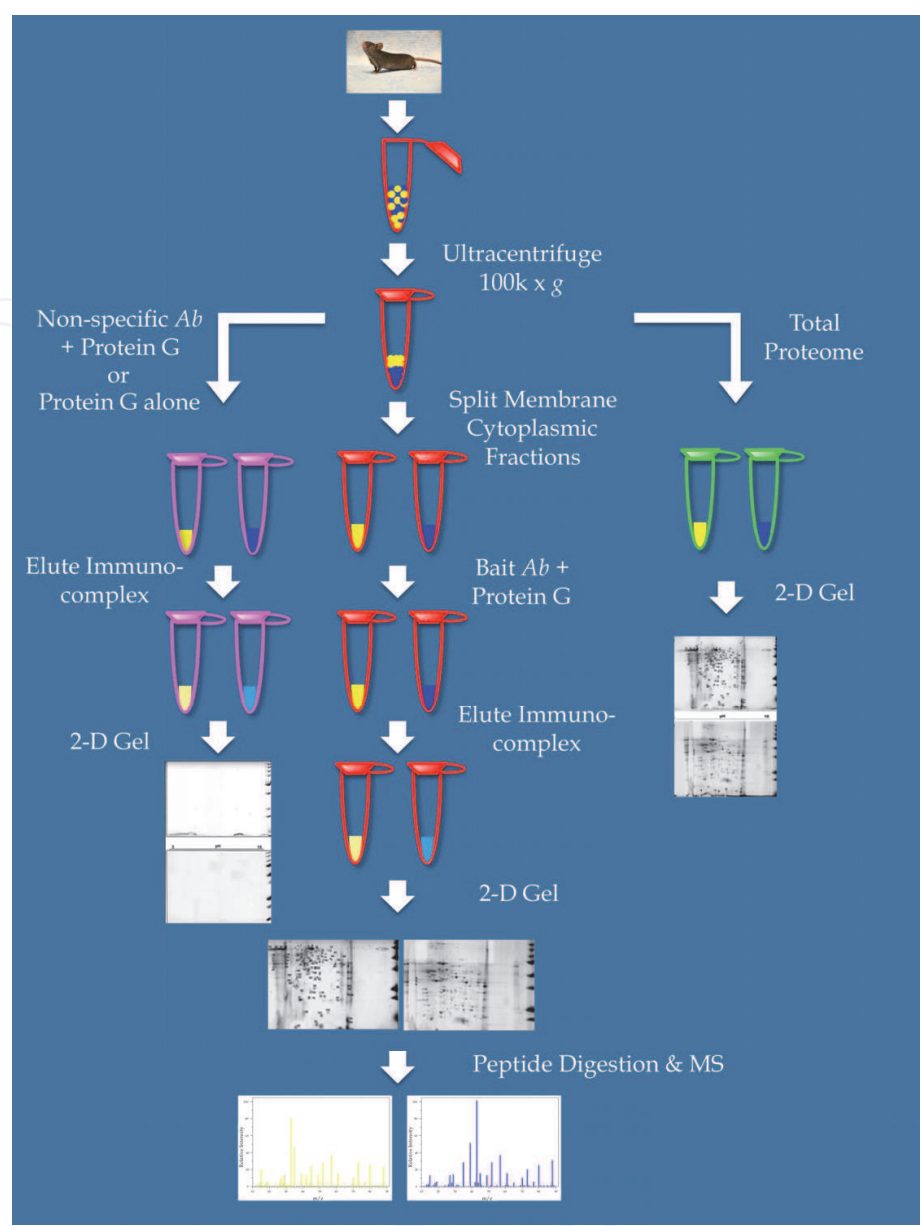
Fig. 2. Schematic of a high-throughput proteome experiment using coimmunoprecipitation, two-dimensional gels, and mass spectrometry. Initially, proteins are solubilized and separated by ultracentrifugation into membrane and cytoplasmic fractions (blue and yellow), which then are divided into two separate aliquots. Anti-bait antibody (*Ab*) with protein G beads (red tubes) is used to coimmunoprecipitate putative protein partners obtained from an organism or tissue lysate. The different subcellular fractions are probed with an antibody to a specific protein and the immunocomplex captured with Protein G-coated beads. The resultant immunocomplexes are eluted, fractionated on two-dimensional gels, and analyzed using LC-MS/MS. Control samples consist of running membrane and cytoplasmic fractions: in the absence of antibody and beads (total proteome; green tubes); with beads alone (purple tubes); or with a nonspecific antibody and beads (purple tubes).

approach to preclearing, but which can also be used as a negative control, is to use a non-specific antibody. The antibody must be isotype specific, when using a monoclonal antibody, or source specific, when using a polyclonal antibody for coimmunoprecipitation.

For example, in the event that the coimmunoprecipitating antibody is a mouse monoclonal IgG1, then use a nonspecific mouse monoclonal IgG1. If, on the other hand, the antibody to be used is a rabbit polyclonal, use a non-specific rabbit polyclonal antibody. When used as a control, mix the precleared lysate with the non-specific antibody and beads and process for gel fractionation. Finally, empirically determine how much antibody to add to the lysate fraction or to the beads, by determining the signal to noise in your result. A good starting point is to begin with 5 μg of antibody and work up or down in concentration from there.

The cells are now ready for coimmunoprecipitation for high-throughput analyses using 2-D gels (**Figure 2**). Control experiments consist of: any sticky proteins adhering to the beads (non-immunocomplexed protein), any proteins obtained using a non-specific antibody, and finally all proteins from the entire proteome of the tissue/organism. While controls may take more samples, they are of value for comparison purposes and troubleshooting, and for acceptance into high impact journals. The gel showing the total proteome is important, since it will provide an overall pattern of protein spots with which to compare the gel containing the immunocomplexed proteins. You will likely see some similar spot patterns between the gels if the separation is of a good quality. One question that will arise is whether to first bind the antibody to the beads or bind the antibody to the antigen in the lysate and then to the beads. One argument for binding the antibody to the beads first is that, since the beads are already covered with antibody, there will be a decrease in contaminant binding, and thus, a decrease in background. Antibody can be covalently bound to beads using Dimethyl Pimelimidate (DMP) Disuccinimidyl Glutarate (DSG), Disuccinimidyl Suberate (DSS), or Disuccinimidyl Tartrate (DST). Arguments against crosslinking include, the buildup of aggregates, antibodies such as monoclonals may lose their affinity, or the antibody cross-links to the beads in the incorrect position causing hindrance to antigen binding.

3.4 Fractionation and gel staining

Once both immunocomplexed and non-immunocomplexed beads are washed, perform the elution step using equal volumes of IEF sample and elution buffers, since the proteins are fractionated in two dimensions. At this point there are various nuances in terms of technique for running a 2-D gel. Among these is a step-by-step description by Kathiresan et al., (2009). Here, we will suggest some of the initial troubleshooting that may be necessary before and/or after running a full-fledged experiment. If little is known about the proteome of the tissue in your experiment, it will be of value to use an Immobilized pH Gradient (IPG) strip with a broad pH range (e.g., pH 3 - 10). Moreover, rather than initially using strips of 18 to 24 cm, which give a better resolution, use a 7 cm strip to get a quick representation of the pI ranges that you will be working with. Also, remember that the protein volume you can load is related to the size of the IPG strip, so that 7, 11, 17, 18, and 24 cm strips require volumes of 125, 185, 300, 325, and 450 μL, respectively. After separation in the first dimension, proteins are fractionated according to weight in the second dimension at which point the gel is prepared for staining.

There are several staining methods available, assuming that a CyDye was not used, since this step is accomplished prior to running the gel. The choice of stain is dependent on whether you are interested in searching for the low hanging fruit, that is, proteins that are highly expressed, or you may wish to increase staining sensitivity to detect as many protein partners as possible. Colorimetric stains include Coomassie Brilliant Blue, which will suffice for the former choice, since this stain will detect in-gel protein concentrations as low as 10

ng. For more inclusive resolution of proteins you can use silver staining, which detects protein concentrations less than 0.25 ng. For a detection range that lies between these two stains, fluorescent dyes are available that detect 0.25 – 2 ng. However, all the stains have their advantages and limitations. Coomassie Blue has less sensitivity, but is probably the most compatible stain for mass spectrometry. Silver staining has greater sensitivity but is less compatible with mass spectrometry, because, as with Coomassie Blue, the protein must be destained prior to tryptic digestion. Since formaldehyde is part of the silver staining process, cross-linking of the protein occurs (Richert et al., 2004), thereby causing problems with protein extraction from the gel and interference with mass spectrometry. A few techniques have been suggested to circumvent this problem, including ammoniacal silver staining (Richert et al., 2004; Chevallet et al., 2006). Moreover, some vendors (e.g., Thermo Scientific Pierce) optimize their reagents to make silver staining more compatible for mass spectrometry. However, silver staining still remains problematic, with its poor reproducibility and a nonlinear dynamic range, when measuring staining intensity relative to the amount of protein. There are many fluorescent stains, including those of a non-covalent variety such as SYPRO Orange, Red, Ruby, and Tangerine (Invitrogen, Carlsbad, CA, USA), ruthenium II, Deep Purple (GE Healthcare, Piscataway, NJ, USA), Krypton (Thermo Scientific, Inc., Rockford, IL, USA), and Oriole (Biorad, Hercules, CA, USA). The advantages are sensitivity, a greater dynamic range, and compatibility with mass spectrometry. Disadvantages lie primarily with cost, because of the necessity for extensive hardware for detection and quantification, the loss of signal with exposure to light, and the potential for masking certain peptides. For example, Deep Purple and SYPRO Ruby begin to lose their fluorescence after two minutes of exposure to UV transillumination, so that by 19 minutes they have lost 83% and 44% of their fluorescence, respectively (Smejkal et al., 2004). SYPRO Ruby may also inhibit identification of cysteine- and tryptophan-containing peptides (Ball & Karuso, 2007). In addition, not all fluorescent stains are compatible with the various gel types that are used to fractionate proteins for LC-MS/MS. Ruthenium II, which is much cheaper than SYPRO stains, causes increased background staining in Bis-Tris gels relative to Tris-Glycine gels (Moebius et al., 2007). These are all factors to keep in mind as part of your experimental design. Finally, once the gel is stained, any vertical streaking of protein is likely the result of insufficient equilibration or problems with the buffer solution, whereas horizontal streaking may be the result of incomplete solubilization, impurities, improper detergent, or an isoelectric focusing time that is too long or too short. With the completion of staining, the gel will need to be destained prior to removal of gel spots either manually or with a robotic arm.

4. Verification of protein partners

4.1 Manual verification of peptides

While the specifics of understanding how the data are obtained and analyzed are beyond the scope of this chapter, a quick review of some highlights are useful before describing potential experiments to validate your interactions. Once tandem mass spectrometry is completed you will obtain data derived from database search engines such as MASCOT, Seaquest, and X! Tandem. Search engines such as MASCOT generate scores as well as a compilation of spectral data. Scores above 60 can be considered valid for protein identification, assuming other parameters such as spectral data are in order. However, one may want to be conservative, especially when examining potential new partners.

Regardless, scores should not be taken at face value without analyzing the fragmentation spectra for each identified peptide, since you can have good scores and bad spectral data as well as bad scores and good spectral data. Personnel from your mass spectrometry core facility can assist in these analyses, however, you should familiarize yourself with how the ion spectra are generated and identified for your own understanding. A good starting point in comprehending spectral data is a tutorial at proteome software.com that comes in the form of a short presentation, http://www.proteomesoftware.com /Proteome_ software_pro_protein_ id.html. In addition, there are many other sources of value in understanding the mechanics of mass spectrometry-based proteomics, including light reading in review articles (Aebersold and Mann, 2003), as well as more intense reading in specialized books (Gross, 2004). Also, before deciding on a core facility to analyze your precious data, assuming you have a choice, you may want to give this sobering article a quick read (Bell et al., 2009). This paper will likely push your choice towards a facility where the personnel have a great amount of experience and the search engines are continually maintained and up-to-date. Finally, the data should be analyzed for false positive rates, since some proteomics journals now require these analyses, including *Molecular and Cellular Proteomics*, which published standards in 2005 (Bradshaw, 2005).

4.2 Reciprocal coimmunoprecipitation

Once you have obtained the results, showing the putative protein-protein interactions, you can use various means to begin assessing their validity using bioinformatics and different experimental procedures. Here, we will discuss some of the methods for experimentally verifying interactions as well as assessing the potential functions of these interactions. The methods to verify are many and can depend on the protein of interest as well as the experimental question. However, one of the first and relatively easy steps for verification, considering that you've just run a two-dimensional gel, is to perform a reciprocal coimmunoprecipitation. Here, the goal is to coimmunoprecipitate the protein that was originally used as bait (antigen) by using the newfound prey (protein partner). In essence, there is a role reversal, so that the former bait is now prey and vice-versa. The means to accomplish this procedure are similar to those used for a coimmunoprecipitation using western blotting. Once you step into the realm of probing the results of a coimmunoprecipitation experiment with an antibody in a western blot, you have to consider the IgG artifacts that may appear on your film. These artifacts are the result of the presence of light (~25 kDa) and heavy (~50 kDa) immunoglobulin fragments that are detected when using an antibody from the same species for both the immunoprecipitation /coimmunoprecipitation pull-down and the western blot. The secondary antibody will recognize both IgG chains, since these are eluted from the beads along with the antibody and fractionated on the gel. The consequence is that the antigen will be masked if it has a weight similar to either IgG. Moreover, if monoclonals are used to both pull down and probe the blot, the secondary recognizes the 25 kDa band; if polyclonals are used for both, the 50 kDa band will appear as artifact. A problematic example is the use of a rabbit polyclonal antibody for both the immunoprecipitation and the western blot. To circumvent this issue, you can use antibodies from different species or use a secondary antibody consisting of HRP bound to Protein A or G. An HRP conjugated to either of these proteins will detect their non-denatured forms but not their denatured forms (Lal et al., 2005). A third solution is to use secondary antibodies that only recognize the light or heavy chain

IgGs (Jackson Immunoresearch Laboratories, Inc., Westgrove PA, USA). Thus, if your protein of interest lies in the 45 - 55 kDa range, you would use a secondary that recognizes only the light chain IgG and vice versa, if your protein lies in the 20 to 30 kDa range. However, one assumption that should not be made in running the reciprocal coimmunoprecipitation is that all the parameters in washing and stringencies are the same as for the original coimmunoprecipitation. We find that at times these variables have to be tweaked slightly differently. However, with practice these issues are usually fixed relatively quickly.

4.3 RNAi and overexpression

A method of verification that can clarify the function of your newly discovered interactions is the use of RNAi in a heterologous expression system. This approach is especially useful for proteins that lend themselves well to this sort of system, such as ion channels. There are several different types of cells to use, including HEK 293 and CHO cells, and if cell polarity is of concern, Madin-Darby Canine Kidney (MDCK) epithelial cells. These expression systems provide a vehicle for, not only expressing your proteins, but also as a means to silence the protein endogenously.
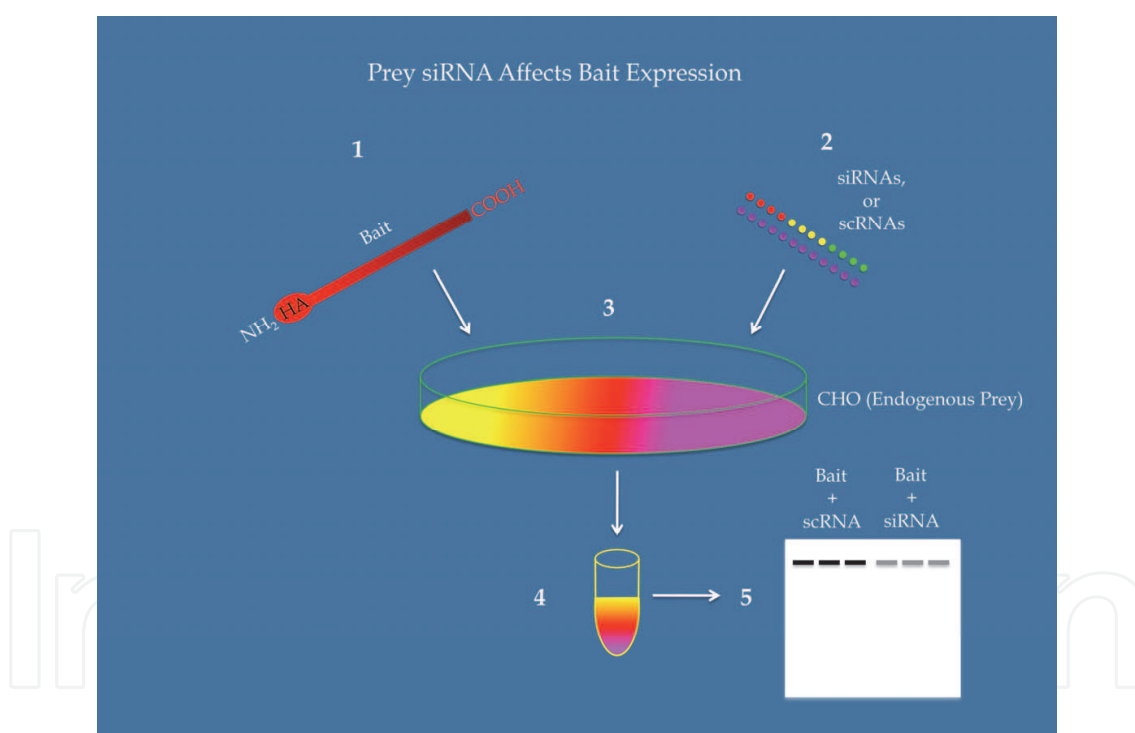

Fig. 3. Experiments, using cDNAs, are conducted to probe the function of protein-protein interactions when the prey protein is silenced in a heterologous expression system. (1) The original bait protein, in the form of an HA-tagged cDNA, is transfected with (2) siRNAs targeting specific nucleotides of the newly- discovered protein partner, or with scrambled RNAs (scRNAs; control) containing nonsense sequences. In this scenario, the (3) prey protein is endogenous to the expression system. Plates for each treatment are prepared in triplicate. (4) Cells from each treatment are scraped, homogenized and prepared for western blotting. (5) The resultant blot is probed for expression of the HA-tagged protein in both treatment and control conditions.

In order to accomplish these experiments, you will need the gene or genes that encode one or both of your proteins, that is, the bait and the prey. If the cDNA comes from either a private source or a vendor, do not assume that the construct you receive contains cDNA with a correct sequence. Be sure that it was sequenced *very* recently before use, as in, after the last amplification, because it is not uncommon to obtain constructs from either source only to discover a mutation during the course of your study, when it's too late. Once sequenced, tag the cDNA on either the C- or N-terminus with one of the specified tags discussed previously. We typically find that the antibodies to the HA and FLAG tags work quite well in these experiments. However, you have to keep in mind that the tag itself may interfere with the interaction. If the coimmunoprecipitation fails, there is no need to panic, because you can just place the tag at the other end. Also, remember that if the tag is placed on the N-terminal end of a signaling protein, you may lose your tag. Inserting the tag farther into the construct resolves this issue.

The decision then comes down to, when to silence and when to over-express, once all the necessary cDNA constructs are in order. In our work, we find that inserting and over-expressing the cDNA, encoding the original bait protein, and silencing the partner (prey) with siRNAs endogenously, works the best in our RNAi experiments (**Figure 3**). This approach entails determining that the partner, pulled from the high-throughput study, is expressed endogenously in your heterologous system. If so, check that the sequences of the siRNAs match those of the endogenous protein found in the heterologous cells. Be sure to have at least three to four siRNAs, targeting 18 – 23 bases of the sequence in different regions that are approximately 70 – 100 bases from either the 5′ or 3′ ends. Search for AA dinucleotides, since siRNAs with an overhanging UU pair at the 3′ end is the most effective, although other dinucleotides are effective to some extent (Elbashir et al., 2001; Elbashir et al., 2002). Avoid runs of G or C, since these are cut by RNAses. In addition, the GC content should not exceed 30 – 50% because the siRNA becomes less active. BLAST siRNA sequences to avoid knockdown of genes with similar sequences.

Once all materials are ready, transfect the cells with the siRNAs and the over-expressed protein (**Figure 3**). A fluorescent tag such as Cerulean can be used to check for the earliest expression of protein in live cells. We find that proteins are expressed within the first two hours, with transfection reagents such as Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA), which can be removed after four hours of transfection. However, the efficiency of transfection may vary from cell type to cell type, so a test of comparable products is needed to find the most efficient one. Anywhere from one to all of the siRNAs can be added in an equivalent ratio. Following transfection, cells are allowed to grow for approximately 48 hours at which time they are processed for protein quantification. Run triplicate plates along with a negative control, such as cells treated with scrambled RNA and over-expressed protein. Experiments are repeated a total of three times. Band densities are measured, averaged, and analyzed for statistical significance by comparing experimental versus control groups. In order to control for protein loading, perform a protein assay and verify by analyzing a control protein on the same blot as experimental and control treatments (e.g., β-actin, GAPDH).

In a similar manner, over-expression of the partner can be managed through transfection of heterologous expression systems (**Figure 4**). Again, the bait protein can be measured, but this time in response to over-expressing the prey as opposed to silencing. Transfect both constructs in a 1:1 ratio and use a control, consisting of empty vector or vector with the construct in reverse sequence along with the construct carrying the prey sequence. This procedure will clarify if the addition of another vector dilutes the expression of the prey protein. Densitometry measurements are made as before and analyzed statistically.

In summary, the search for mechanisms that regulate the many proteins in different cell systems can be tackled using a variety of different techniques. The results can provide you with a bounty of data that can be verified and used to mine many different databases. The outcomes from these experiments will provide you with new and fascinating insights that, heretofore, you may not have thought about. The critical issues are that you will need to obtain a clear understanding of what is occurring to the proteins at different stages of the experiment. This understanding will allow you to obtain clean representations of the proteins to be assessed with fewer inherent artifacts. Once mass spectrometry is completed, the data can then be used to mine various databases in order to fit and expand your data into an interactome.
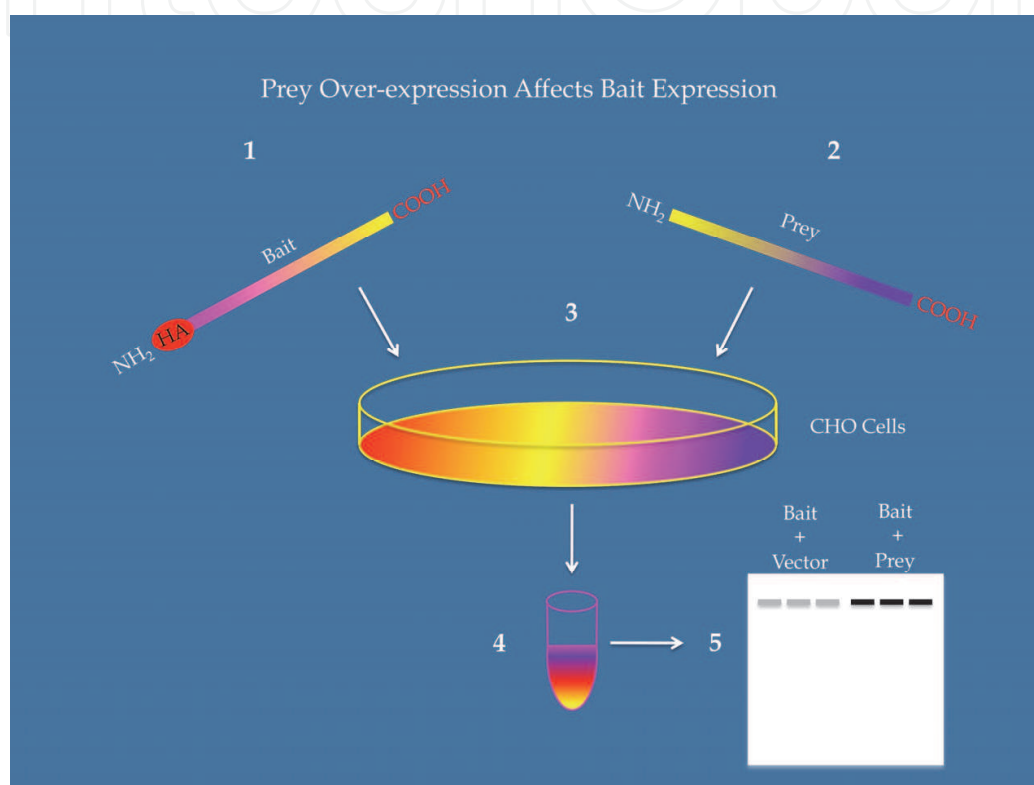
Fig. 4. Experiments, using cDNAs, are conducted to probe the function of protein-protein interactions when the prey is over-expressed in a heterologous expression system. **(1)** The original bait with the **(2)** newly-discovered prey or the bait with plasmid containing the reverse sequence of the prey (control) are cotransfected into **(3)** CHO cells. **(4)** Cells are collected and prepared for western blotting, where expression of the bait protein is probed **(5)** for both treatment and control conditions.

5. Bioinformatics

5.1 Comparison of high-throughput data with existing data

Once the initial experiments are completed, the lab worker will often wish to analyze the data by comparison to known interactions already existing in molecular interaction databases. These resources exist to collate and curate experimental data from laboratories around the world. Initially, interaction databases were established in isolation and often performed redundant curation of the same, high-visibility papers to their own standards, subsequently releasing the data in their own proprietary formats. However, more interaction data is

published in any one calendar month than can be captured by these resources collectively, and no single database can claim to have complete coverage of the literature. In order to approach a complete interactome for the organism or process of interest, the user has always had to combine data from multiple resources. This was well nigh impossible prior to 2004, as the different data formats, required separate parsers to be written for each data source. This approach began to change with the release of the first HUPO-PSI standard representation of interaction data. Nowadays, two related formats exist – PSI-MI XML2.5 and MITAB2.5 (Kerrien et al., 2007), which are supported by both the majority of interaction databases and also by related visualization and analytical resources. These formats have enabled consistent data capture by multiple resources, with the choice of XML or tab-deliminated files often driven by either the complexity of data that the user wishes to harvest, or the amount of bioinformatic support available to them. Controlled vocabularies now make the terminology used to annotate these data consistent across the many data resources.

One major advantage of multiple resources sharing the same data format is that it is now possible to simultaneously access multiple resources with a single query, cluster the results, and visualize these in a single graph. A PSI Common Query InterfaCe (PSICQUIC) was developed that allows software clients to interact with multiple services and is based on the existing PSI MI file formats and the new Molecular Interaction Query Language (MIQL) (Aranda et al., in preparation). MIQL is based on standard Lucene syntax (http://lucene.apache.org/) and offers single word or phrase queries (abl1 AND "pull down"), search in specific data attributes/columns (abl1 AND species:human), wildcards (abl*), and logical operators. At the time of writing, 16 data providers were providing PSICQUIC servers, with a total of 16 million interactions available to query. PSICQUIC lays no constraint on data type or quality and much of what is available is also redundant, in that databases, which do not have their own curation team, will import information from those that do. To address this issue, several of the major databases have come together to synchronize their curation rules and data release through the IMEx Consortium (www.imexconsortium.org). This consortium allows the user to access and download a non-redundant, consistently annotated set of data, again using a PSICQUIC client to access appropriately tagged records (Orchard et al. in preparation).

5.2 Data resources

A number of databases exist, many of which have a bias in their curation strategy, either towards particular organisms or cellular processes. A brief summary of a number of these resources is given below, a more complete but less detailed list can be obtained from Pathguide (www.pathguide.org/). All the databases listed make their data available both through a dedicated website and also from their respective ftp sites in one or both of the PSI formats.

IntAct (www.ebi.ac.uk/intact) – no species or process bias, collects data from all organisms. Mainly contains protein-protein interactions but also annotates protein-small molecule, protein-nucleic acid. Interactions are derived from literature curation or direct user submissions and are freely available. Database and associated tools are open-source and available for download. IntAct provides a PSICQUIC service and is a full member of IMEx.

MINT (http://mint.bio.uniroma2.it/mint/) - no species or process bias, collects data from all organisms. Focuses on experimentally verified protein-protein interactions mined from the scientific literature. MINT provides a PSICQUIC service and is a full member of IMEx.

DIP (http://dip.doe-mbi.ucla.edu/dip) - no species or process bias, collects data from all organisms. Catalogs experimentally determined protein-protein interactions between proteins. The data are both manually curated and also automatically, using computational approaches that utilize knowledge about the protein-protein interaction networks extracted from the most reliable, core subset of the DIP data. DIP provides a PSICQUIC service and is a full member of IMEx.

MatrixDB (http://matrixdb.ibcp.fr) - focuses on interactions established by extracellular proteins and polysaccharides, mainly in human and mouse. Reports interactions of proteins, individual polypeptide chains or multimers (permanent complexes) and carbohydrates. MatrixDB provides a PSICQUIC service and is a full member of IMEx.

InnateDB (www.innatedb.ca/)- a database of the genes, proteins, experimentally-verified interactions and signaling pathways involved in the innate immune response of humans and mice to microbial infection. Integrates known interactions and pathways from major public databases together with manually-curated data. InnateDB provides a PSICQUIC service and is a full member of IMEx.

MPIDB (http://jcvi.org/mpidb/) - collects physical microbial interactions manually curated from the literature or imported from other databases. MPIDB provides a PSICQUIC service and is a full member of IMEx.

BioGrid (http://thebiogrid.org/) – a resource of protein–protein and genetic interactions for many model organism species. BioGRID provides a PSICQUIC service and is an observer member of IMEx.

HPRD (www.hprd.org/) – a database of human interactions, mixed species interactions (e.g. human-mouse) are modeled to human. Commercial entities have to pay a fee to use the data, under a licensing arrangement.

5.3 Data visualization

Once the user has downloaded their required data from one or all of the listed resources, the next step is to combine the networks then import and overlay your own data. The tool of choice for this exercise is most commonly Cytoscape (www.cytoscape.org), a free software package for visualizing, modeling and analyzing molecular and genetic interaction networks. In Cytoscape, nodes representing biological entities, such as proteins or genes, are connected with edges representing pairwise interactions, such as experimentally determined protein–protein interactions. Nodes and edges can have associated data attributes describing properties of the protein or interaction. Cytoscape allows users to extend its functionality by creating or downloading additional software modules known as 'plugins'. These plugins provide additional functionality in areas such as network data query, network data integration and filtering, attribute-directed network layout, Gene Ontology (GO) enrichment analysis and network motif functional module, protein complex or domain interaction detection.

Network data can be imported into Cytoscape is several formats, including those of the PSI-MI. To merge files in Cytoscape, the gene or protein identifier in the file must exactly match the corresponding Cytoscape node ID (or other Cytoscape attribute that has been previously loaded). If no matching identifiers are present, the situation can be corrected by loading an additional identifier into Cytoscape as a new node attribute. The ID mapping service supplied by UniProt (www.uniprot.org/) or by PICR (www.ebi.ac.uk/Tools/picr/) are recommended to achieve this correction. The Advanced Network Merge, a core Cytoscape plugin will allow network merging on the click of a single button.
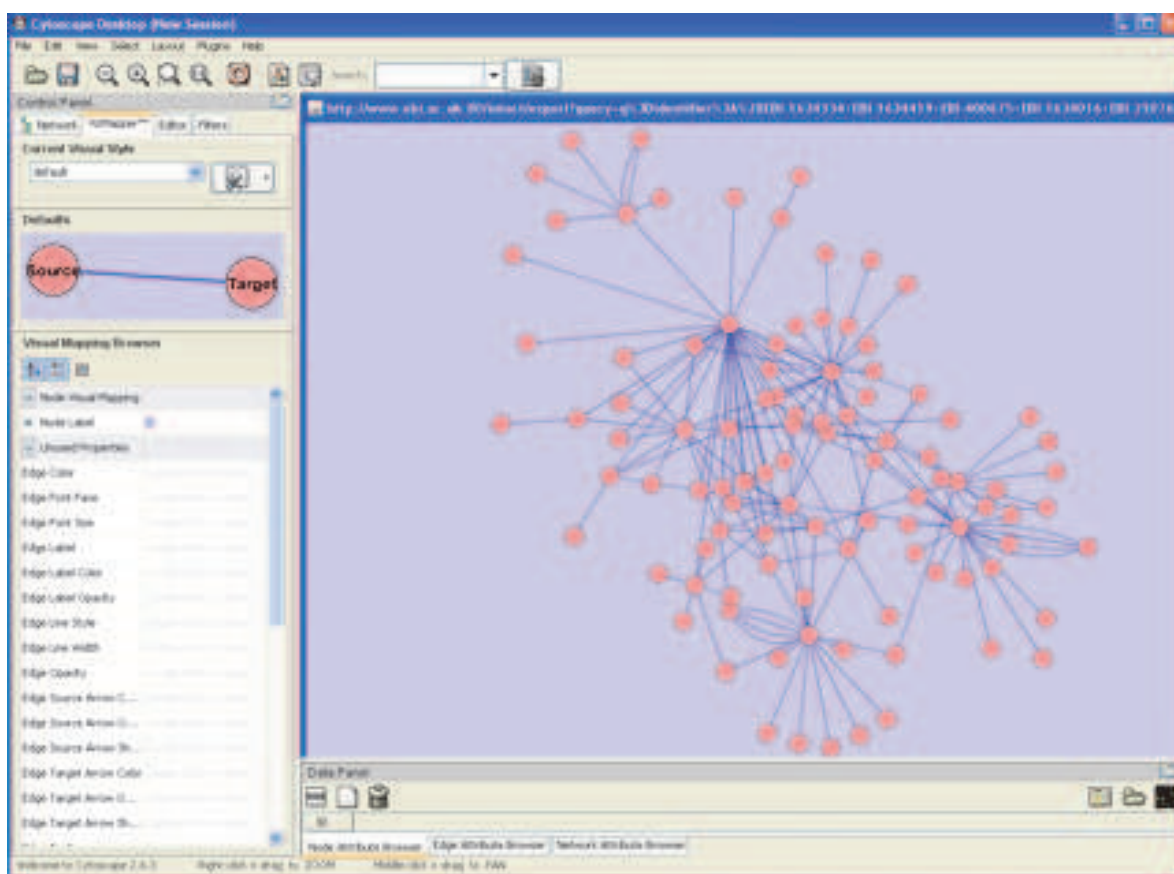
Fig. 5. An interaction network as viewed in Cytoscape.

5.4 Interolog mapping

Large-scale PPI networks are only available for a limited number of model organisms, therefore, groups working on less well-studied organisms have to rely on network inference using the interolog concept originally introduced by Walhout et al. (2000). This concept combines known PPIs from one or more source species and orthology relationships between the source and target species to predict PPIs in the target species. There are a number of resources available which perform the orthology mapping Inparanoid (http://inparanoid.sbc. su.se/cgi-bin/index.cgi) and Compara (www.ensembl.org /info/docs/api/ compara/index.html) being probably the best known. Few tools exist for interolog mapping, however, two database resources exist in which this exercise has been pre-computed for the user; STRING (http://string-db.org/) transfers associations/interactions between several hundred organisms and InteroPORC (http://biodev.extra.cea.fr/interoporc/) for the fully sequenced organisms described in the Integr8 database (www.ebi.ac.uk/integr8). Both resources make the data available in PSI format and both have a PSICQUIC server. Additionally, the InteroPORC software is freely available for in-house use (Michaut et al. 2008).

6. Conclusions

The data generated from experiments that examine genes and proteins has increased logarithmically over the last 20 years, largely driven by recent advances in high-throughput technologies that examine proteins individually, as well as in complexes. High-throughput

protein studies that combine coimmunoprecipitations with 2-D gels have increased as a result of the higher quality of data obtained from mass spectrometry. The advent of these technologies has helped to fuel a need for the formation of many curated databases, such as those that capture molecular interactions. These molecular interactions databases are increasing their usefulness to the community by making their datasets available in a single, unified format. In addition, many are also linked in a unifying organization, such as the IMEx Consortium, which is ensuring that the user can download a non-redundant set of consistently annotated data. This means that the user now has a single point of entry from which to download data, and an increasing number of tools with which to subsequently analyze those data.

7. Acknowledgements

8. References

Andrews, C.B., & Gregory, T.R. (2009) Genome Size is Inversely Correlated with Relative Brain Size in Parrots and Cockatoos. *Genome*, Vol.52, No.3 (March 2009), pp. 261–267, ISSN 0831-2796.

Ball, M.S., & Karuso P. (2007) Mass Spectral Compatibility of Four Proteomics Stains. *Journal of Proteome Research*, Vo.6, No.11 (November 2007), pp. 4313-4320, ISSN 1535-3893.

Bell, A.W., Deutsch, E.W., Au, C.E., Kearney, R.E., Beavis, R., Sechi, S., Nilsson, T., & Bergeron, J.J. & HUPO Test Sample Working Group (2009) A HUPO Test Sample Study Reveals Common Problems in Mass Spectrometry-Based Proteomics. *Nature Methods*, Vol.6, No.6, (June 2009), pp. 423-430, ISSN 1548-709.

Bradshaw RA. (2005) Revised Draft Guidelines for Proteomic Data Publication. *Molecular and Cellular Proteomics,* Vol.4, No.9, pp. 1223-1225, ISSN 1535-9476.

Chevallet, M., Diemer, H., Luche, S., van Dorsselaer, A., Rabilloud, T., & Leize-Wagner, E. (2006) Improved Mass Spectrometry Compatibility is Afforded by Ammoniacal Silver Staining. *Proteomics*, 2006, Vol.6, No.8 (April 2006), pp. 2350–2354, ISSN 1615-9853.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A.R., Vailaya, A., Wang, P.L., Adler, A., Conklin, B.R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G.J., Ideker, T., Bader, G.D. (2007), Integration of Biological Networks and Gene Expression Data Using Cytoscape. *Nature Protocols*, Vol. 2, No. 10 (September 2007), pp. 2366-2382, ISSN 1754-2189.

Duzhyy, D., Harvey, M., & Sokolowski, B. (2005) A Secretory-Type Protein, Containing a Pentraxin Domain, Interacts With an A-type K+ Channel. *Journal of Biological Chemistry*, Vol.280, No.15, (April 2005), pp. 15165-15172, ISSN 0021-9258.

Elbashir, S.M., Harborth, J., Weber, K., & Tuschl, T. (2002) Analysis of Gene Function in Somatic Mammalian Cells Using Small Interfering RNAs. *Methods*, Vol.26, No.2, (February 2002), pp. 199-213. ISSN 1046-2023.

Elbashir, S.M. Martinez, J., Patkaniowska, A., Lendeckel, W., & Tuschl, T. (2001) Functional Anatomy of siRNA for Mediating Efficient RNAi in *Drosophila melanogaster* Embryo

Lysate. *European Molecular Biology Organization Journal*, Vol.20, No.23, (December 2001), pp. 6877-6888, ISSN 0261-4189.

Eng, J.K., McCormack, A.L., & Yates III, J.R. (1994) An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *Journal of the American Society Mass Spectrometry*, Vol.5, pp. 976–989. ISSN 1044-0305.

Evan, G.I., Lewis, G.K., Ramsa, G., & Bishop, J.M. (1985) Isolation of Monoclonal Antibodies Specific for Human c-myc Proto-onco- Gene Product. *Molecular and Cellular Biology*, Vol.5 No.12, (December 1985), pp. 3610–3616. ISSN 0270-7306.

Fox, A.D., Hescott, B.J., Blumer, A., & Slonim, D.K. (2011) Connectedness of PPI Network Neighborhoods Identifies Regulatory Hub Proteins. *Bioinformatics*, Vol.27, No.8, (April 2011), pp. 1135-1142,, ISSN 1367-4803.

Gross J.H. (2004) Mass Spectrometry. Springer-Verlag, New York, NY 592 pp, ISBN 3540407391.

Golemis, E.A., & Adams, P.D. (2005) Protein-Protein Interactions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 938 pp, ISBN 0-87969-723-7.

Harvey, M.C., & Sokolowski, B.H.A. (2009) In Vivo Verification of Protein Interactions in the Inner Ear by Coimmunoprecipitation. In: *Auditory and Vestibular Research, Methods and Protocols,* B. Sokolowski (Ed.), Vol.493, No.II, 299-310, Humana Press, Springer Protocols. New York, NY, ISBN 978-1-934115-62-6.

Hopp, T.P., Prickett, K.S., Price V.L., Libby, R.T., March, C.J., Cerett, D.P., Urdal, D.L., & Conlon, P.J. (1988) A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Nature Biotechnology*, Vol.6, (October 1988), pp. 1204–1210, ISSN 1087-0156.

Huber, L.A., Pfaller, K., & Vietor, I. (2003) Organelle Proteomics: Implications for Subcellular Fractionation in Proteomics. *Circulation Research*, Vol.92, No.9, (May 2003), pp. 962-968, ISSN 1524-4571.

Kathiresan, T., Harvey, M.C., & Sokolowski, B.H.A. (2009) The Use of 2-D Gels to Identify Novel Protein–Protein Interactions in the Cochlea. In: *Auditory and Vestibular Research, Methods and Protocols,* B. Sokolowski (Ed.), Vol.493, No.II, 269-286, Humana Press, Springer Protocols. New York, NY, ISBN 978-1-934115-62-6.

Kerrien S, Orchard S, Montecchi-Palazzi L, Aranda B, Quinn AF, Vinod N, Bader G, Xenarios I, Wojcik J, Sherman D, Tyers M, Salama JJ, Moore S, Ceol A, Chatr-aryamontri A, Oesterheld M, Stümpflen V, Salwinski L, Nerothin J, Cerami E, Cusick ME, Vidal M, Gilson M, Armstrong J, Woollard P, Hogue C, Eisenberg D, Cesareni G, Apweiler R, Hermjakob H (2007) Broadening the Horizon--Level 2.5 of the HUPO-PSI Format for Molecular Interactions. *BMC Bioinformatics* Vol.5, No.44 (October 2007), ISSN 1471-2105.

Lal, A., Haynes, S.R. & Gorospe, M. (2005) Clean Western Blot Signals From Immunopre-cipitated Samples. *Molecular and Cellular Probes*, Vol.19, No.6, (December 2005), pp. 385–388, ISSN 0890-8508.

Mateo, C., Fernandez-Lorente, G., Pessela, B.C., Vian, A., Carrascosa, A.V., Garcia, J.L., Fernandez-Lafuente, R., & Guisan J.M. (2001) Affinity Chromatography of Polyhistidine Tagged Enzymes. New Dextran-Coated Immobilized Metal Ion Affinity Chromatography Matrices for Prevention of Undesired Multipoint Adsorptions. *Journal of Chromatography A*, Vol.915, No.1-2. (April 2001), pp. 97-106, ISSN 021-9673.

McGrath C.L. &. Katz, L.A. (2004) Genome Diversity in Microbial Eukaryotes. *Trends in Ecology and Evolution,* Vol.19, No.1 (January 2004), pp. 32-38, ISSN 0169-5347.

Michaut, M., Kerrien, S., Montecchi-Palazzi, L., Chauvat, F., Cassier-Chauvat, C., Aude, J.C., Legrain, P., Hermjakob, H. (2008) InteroPORC: Automated Inference of Highly

Conserved Protein Interaction Networks. *Bioinformatics*, Vol. 24, No.14 (July 2008), pp. 1625-1631, ISSN 1367-4803.

Moebius, J., Denker, K., & Sickmann, A. (2007) Ruthenium (II) Tris-Bathophenanthroline Disulfonate Is Well Suitable for Tris-Glycine PAGE But Not for Bis-Tris Gels. *Proteomics* Vol7, No.4, (February 2007), pp. 524–527, ISSN 1615-9853.

Moore, G.E. Cramming More Components Onto Integrated Circuits. *Electronics (Magazine)*, Vol.38, No.8, (April 1965), pp. 114-117, Available from ftp://download.intel.com /museum/Moores_Law/Articles-press_Releases/Gordon_Moore_1965_Article. pdf.

Navaratnam, D., (2009) Yeast Two-Hybrid Screening to Test for Protein-Protein Interactions inf the Auditory system. In: *Auditory and Vestibular Research, Methods and Protocols,* B. Sokolowski (Ed.), Vol.493, No.II, 299-310, Humana Press, Springer Protocols. New York, NY, ISBN 978-1-934115-62-6.

Noronha, S., Kaufman, J., & Shiloach, J. (1999) Streamline Chelating for Capture and Purification of Poly-His-Tagged Recombinant Proteins. *Bioseparation*, Vol.8, (1999), pp. 145–151, ISSN 0923-179X.

Olivier, M., Aggarwal, A., Allen, J., Almendras, A.A., Bajorek, E.S., Beasley, E.M., Brady, S.D., Bushard, J.M., & Stanford Human Genome Group (2001) A High-Resolution Radiation Hybrid map of the Human Genome Draft Sequence. *Science*, Vol.291, No.5507, (February 2001), pp. 1298-1302, ISSN 0036-8075.

Pedersen RA. (1971) DNA Content, Ribosomal Gene Multiplicity, and Cell Size in Fish. *Journal of Experimental Zoology*, Vol.177, No.1, (May 1971), pp. 65-78. ISSN 1932-5231.

Pellicer, J., Fay, M.F., & Leitch, I.J. (2010) The Largest Eukaryotic Genome of Them All? *Botanical Journal of the Linnean Society*, Vol.164, No.1 (September 2010), pp. 10–15, ISSN 0024-4074.

Richert, S., Luche, S., Chevallet, M., Van Dorsselaer, A., Leize-Wagner, E., & Rabilloud, T. (2004) About the Mechanism of Interference of Silver Staining with Peptide Mass Spectrometry. *Proteomics*, Vol.4, No.4 (April 2004), pp. 909–916, ISSN 1615-9853.

Sanger, F., Nicklen, S., & Coulson, A.R. (1977) DNA Sequencing with Chain-Terminating Inhibitors, *Proceedings of the National Academy of Sciences USA*, Vol.74, No.12, (December 1977), pp. 5463–5467, 0027-8424, ISSN 0027-8424.

Smejkal, G.B., Robinson, M.H., & Lazarev, A. (2004) Comparison of Fluorescent Stains: Relative Photostability and Differential Staining of Proteins in Two-Dimensional Gels. *Electrophoresis*, Vol.25, No.15, (August 2004), pp. 2511-2519, ISSN 0173-0835.

Terpe, K. (2003) Overview of Tag Protein Fusions: From Molecular and Biochemical Fundamentals to Commercial Systems. *Applied Microbiology and Biotechnology*, Vol.60, No.5, (January 2003), pp. 523-533, ISSN 0175-7598.

Venter, J.C., Adams, M.D., Myer, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., & Celera Genome Working Group (2001) The Sequence of the Human Genome, *Science*, Vol.291, No.5507, (February 2001), pp. 1304-1351, ISSN 0036-8075.

Vetter, D.E., Basappa, J., & Turcan, S. (2009) Multiplexed Isobaric Tagging Protocols for Quantitative Mass Spectrometry Approaches to Auditory Research, In: *Auditory and Vestibular Research, Methods and Protocols,* B. Sokolowski (Ed.), Vol.493, No.II, 345-366, Humana Press, Springer Protocols. New York, NY, ISBN 978-1-934115-62-6.

Vinogradov, A.E. (2003) Selfish DNA is Maladaptive: Evidence From the Plant Red List. *Trends in Genetics*, Vol.19, No.11, (November 2003), pp. 609–614, ISSN 0168-9525.

Walhout, A.J., Boulton, S.J., Vidal, M. (2000) Yeast Two-Hybrid Systems and Protein Interaction Mapping Projects for Yeast and Worm. *Yeast,* Vol.17, No.2, (June 2000), pp.88-94, ISSN 0749-503X.
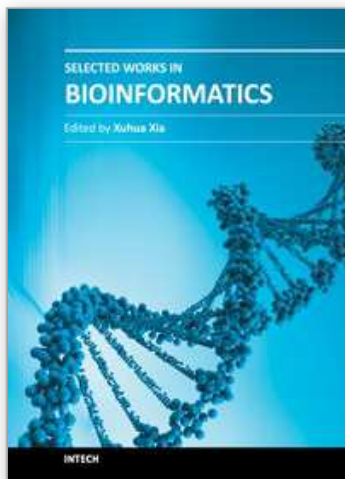
Selected Works in Bioinformatics

Edited by Dr. Xuhua Xia

This book consists of nine chapters covering a variety of bioinformatics subjects, ranging from database resources for protein allergens, unravelling genetic determinants of complex disorders, characterization and prediction of regulatory motifs, computational methods for identifying the best classifiers and key disease genes in large-scale transcriptomic and proteomic experiments, functional characterization of inherently unfolded proteins/regions, protein interaction networks and flexible protein-protein docking. The computational algorithms are in general presented in a way that is accessible to advanced undergraduate students, graduate students and researchers in molecular biology and genetics. The book should also serve as stepping stones for mathematicians, biostatisticians, and computational scientists to cross their academic boundaries into the dynamic and ever-expanding field of bioinformatics.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Bernd Sokolowski and Sandra Orchard (2011). Exploiting Protein Interaction Networks to Unravel Complex Biological Questions, Selected Works in Bioinformatics, Dr. Xuhua Xia (Ed.), ISBN: 978-953-307-281-4, InTech, Available from: http://www.intechopen.com/books/selected-works-in-bioinformatics/exploiting-protein-interaction-networks-to-unravel-complex-biological-questions

INTECH
open science | open minds