Journal of Current Medical Research and Opinion
Received 27-07-2024 | Revised 28-07-2024 | Accepted 04-09-2024 | Published Online 05-09-2024

DOI: https://doi.org/10.52845/CMRO/2024/7-9-4 ISSN (O) 2589-8779 | (P) 2589-8760
CMRO 07 (09), 3531-3548 (2024) www.cmro.in

Original Research

Advanced Technologies for Studies on Protein Interactomes: Protein Microarray, Protein Chips, Yeast Two-Hybrid System, Mass Spectrometry and cDNA Library Screening

Zahraa Wahid Saeed Baha[1] | Sama Mazen Abdul Razzaq Karim[2]

[1]University of Babylon, College of Science, Department of Biotechnology, Iraq

[2]University of Babylon, College of Science, Department of Biotechnology, Iraq

Abstract:

Assigning functions to the numerous genes uncovered by large-scale sequencing projects is a major obstacle for modern biology, as only a small portion of gene activity can be deduced directly from the coding sequence. Understanding the function of proteins relies heavily on identifying their interactions with one another. Direct detection of intermolecular interactions through protein microarray, yeast two-hybrid system, mass spectrometry fluorescent techniques to visualise protein complexes or pull-down assays, and technologies detecting functional interactions between genes, such as RNAi knock down or functional screening of cDNA libraries, are the main technologies used to investigate protein-protein interactions and assign functions to proteins. Significant progress has been achieved in the aforementioned methods in the past few years. Here we take a look back at recent advances and how they've changed gene function annotation. Several ProteinChip arrays with chemically modified surfaces (pre-activated) that form covalent bonds with free amine groups are available from Bio-Rad Laboratories (Hercules, CA). They are developed for diverse biological applications including biomarker discovery, protein profiling, protein–protein interaction studies, peptide mapping for protein ID, immunoassays, and receptor-ligand binding studies. These arrays have eight 2-mm diameter spots (A–H format) that correlate with the spacing of wells in a single column of a standard 96-well microplate and are therefore amenable for high-throughput applications in varied robotics. Here, we describe two types of arrays that are particularly amenable to covalently immobilize biomolecules for subsequent capture of targeted proteins from complex biological samples. The arrays differ in their surface chemistry and hence both should be tested to determine the most suitable array for the application envisioned. Other considerations include lower non-specific binding and higher sensitivity.

Key words: cDNA Library Screening, Protein Microarray, Yeast Two-Hybrid System, Mass Spectrometry.

Introduction:

The Human Genome Programme was one of several successful high-throughput sequencing projects; others included the IMAGE consortia and RIKEN's expressed Sequence Tag (EST) sequencing efforts, which resulted in an unprecedented amount of primary DNA sequences from different species. Finding potential genes that possess a certain trait or perform a specific function is a significant undertaking in the post-sequencing stage. The majority of human genes still have no known biological function, despite tremendous bioinformatics attempts to anticipate gene functions using information on orthologues. Only about a third of the human genes that have been predicted so far have a known function. Consequently, we need to use a mix of methods to fill this knowledge gap on gene function generally and mammalian gene function specifically. The goal of many well-designed and extensively-used experimental approaches is to understand gene function by determining which proteins interact with the target gene product. Identifying the protein's interacting partners and describing its localisation inside the cells might provide a good idea of its function, since the role of a new protein is reflected in its interactions with others. An growing number of proteins with unknown functions are being characterised on a worldwide scale utilising high-throughput techniques, due to the sheer volume of unique sequence information [1, 2]. These types of activities are commonly known as proteomics investigations. The first use of the term "proteome" to characterise a genome's protein complement was in 1995. One branch of proteomics is known as "expression proteomics," which looks at how proteins are expressed on a global scale, and the other is called "cell-map proteomics," which is all about studying complexes of proteins. Worldwide, how do proteins interact and what do they do? These are two of the most pressing problems in post-genomics study. To get closer to these goals, it is crucial to examine protein-protein interactions at a system level, since the majority of pharmacological targets are proteins [3]. Recent breakthroughs in important methodologies exploring protein-protein interactions will be covered in this review.

Protein Chips and Microarrays for Antibodies:

Scientists may now create protein-level maps of cellular networks because to recent developments in high-throughput technologies. Protein microarrays, which are similar to DNA microarrays but serve a different purpose, have become a popular tool for these types of studies. Using antibodies that have been produced against a specific set of target proteins is one use of this array technology. It is possible to bind antigens to their corresponding antibodies by first anchoring these antibodies to a support, like a glass slide, and then running a mixture of protein samples over the surface. Cells treated with various compounds and/or external stimuli of interest have their protein expression levels studied using this method. In their study, Sreekumar et al. (2001) used an antibody array to detect protein expression levels in cancer cells exposed to ionising radiation. They found several new proteins, including DFF40/CAD and CEA, in addition to well-known proteins p53 and DR5, which are regulated by radiation. Haab et al. (2001) provided an additional example by describing the use of a robotic device to detect antigen or antibody solutions on microscope slides. In this investigation, fluorescent labels were applied to the experimental protein samples. This technique was found to have sufficient sensitivity and accuracy in detecting antigen-antibody interactions, making it suitable for use in clinical and research settings to characterise concentrations of several proteins in complicated solutions. Different immobilised probes are used on the surface of other customised protein arrays. A variety of tiny synthetic compounds, polysaccharides [4, 5], allergens, short peptides, or aptamers (oligonucleotide/peptide) can serve as the probes. Review by Zhu et al. (2003) states that protein microarrays are sensitive and practical tools for researching protein-nucleic acid, protein-small molecule, protein-drug, and protein-protein interactions. With the protein microarray,

researchers can analyse the dynamics of protein-protein interactions using real-time detection methods, in addition to discovering new proteins that interact with the target protein.

A Two-Hybrid System for Yeast:

In 1989, Fields and Song used the yeast Saccharomyces cerevisiae transcriptional activator GAL4 to create the yeast two-hybrid method, which was initially used to explore protein-protein interactions. The method employed in this research relies on the GAL4 protein, which has two separate but complementary domains: the binding domain (BD) at the protein's N-terminus, which binds to particular DNA sequences, and the activator domain (AD) at its C-terminus, which contains acidic regions required to activate transcription. The result was a system consisting of two hybrid proteins: one with the GAL4 DNA-binding domain (BD) linked to a protein "X" (bait) and the other with the GAL4 activating region (AD) attached to a protein "Y" (prey). Transcription of the reporter gene controlled by GAL4 can be begun if proteins X and Y can form a protein-protein complex that brings the two domains of GAL4 into proximity [6]. Scientists have refined the time-honored method of using a two-hybrid system over the years. Although these variations are based on the same basic ideas, they have been "fine-tuned" for different applications thanks to the changes made. The protein binds to AD in a one-hybrid method that was designed to explore protein-DNA interaction. The reporter gene is cloned upstream of the DNA segment of interest. When the protein connects to the target DNA (the promoter), the reporter gene's transcription starts. When trying to understand how proteins work, it can be helpful to introduce point mutations, deletions, or inhibitors—either protein or pharmacological—in order to disrupt specific interactions. As a result, two-hybrid systems that reverse and counter-select were created. In these setups, cells perish because the wild-type protein, which can bind to the bait, makes them vulnerable to selection reagents (such as cytotoxic chemicals). Cells that express proteins with mutations that impair the connection are the only ones that will make it. Thus, new prey mutants that are unable to connect to the target protein (the bait) can be found using this method. Similarly, entities with the ability to block particular protein-protein interactions can also be identified by screening chemical compound libraries. For a long time, the fact that yeast lacks the modifying enzymes necessary to post-translationally modify proteins—particularly those involved in signal transduction pathways in higher eukaryotic organisms—was a big drawback of the classic two-hybrid system. To guarantee that prey proteins are functional in the host, the three-hybrid method was created and used to co-transfect the essential enzyme, which allowed for post-translational modification of the prey proteins. Using this method, the changes needed to map a specific inter-protein interaction can be quickly identified. Protein phosphorylation is the most prevalent of these changes; it is typically triggered by certain stimuli and necessitates the production of specific kinases. Osborne et al. published a library screening study that used the Y3H strategy to find proteins that interact with an IgE receptor-derived, phosphorylated bait that contains immunoreceptor tyrosine-based activation motifs (ITAMs). In the bait, a new member of the SH2-containing [7, 8] family was isolated using the gamma subunit of the high-affinity IgE receptor, FcεRI. This member interacts with the Syk or Lyn SH2 domains, which in turn requires the phosphorylation of the ITAMs by tyrosine kinases. Along with the "bait" and "prey" came a plasmid that encodes the tyrosine kinase. Although this approach was initially developed to study tyrosine phosphorylation-mediated interactions, it can be modified to study other post-translational changes as well. Detecting weak interactions between many proteins is another application of the three-hybrid method. Proteins typically form huge multi-component complexes with weak and strong interactions after binding to several additional proteins. It is possible to find new proteins that interact poorly with an existing protein by co-

expressing a known interacting protein; this may serve as a bridge, increasing the affinity of the connection between the two proteins.

Some hosts, like E. coli, are not part of the traditional Saccharomyces cerevisiae system. also been utilised. Faster growth, better transformation efficiency, elimination of the need for nuclear localisation, and the fact that domains with eukaryotic activation domains do not activate E were among the many alleged benefits over the yeast system. transcribing E. coli, and less connections that are indirect and involve endogenous proteins as bridges. One study that took use of the E. To find proteins that interact with each other in the Tat route, the E. coli two-hybrid method was developed. The two reporter systems that utilised the Tat route were an enzymatic assay with a chromogenic substrate and a growth-on-selective-media (maltose-based) approach. In contrast to previous research [9], this E. coli two-hybrid system enhanced the precision of two-hybrid investigations that encompassed the entire proteome.

The initial two-hybrid setup required nuclear translocation of interacting fusion proteins (bait and prey) to initiate reporter gene transcription. As a result, the system is unable to detect as many interactions. For example, misfolding and nuclear localisation issues make full-length trans-membrane proteins unsuitable as baits. To get over this problem, researchers came up with ways to make the two-hybrid system work in both the cytoplasm and the membrane. Instead of building two fusion proteins using nuclear transcription factors, β-galactosidase was separated into two parts and put back together through a bait/prey interaction. The activity of β-galactosidase was used to measure how strong the relationship was between the bait and prey. This allowed researchers to examine the protein-protein interaction in its native subcellular environment .

An alternative method made use of the Ras-controlled signalling cascade on the plasma membrane of Saccharomyces cerevisiae, which contains the temperature-sensitive Ras Guanine Exchange factors (GEF) Cdc25-2. A change from Ras's inactive GDP-bound state to its active GTP-bound state is induced by GEFs. Among the six H proteins, sos is one. Qian • E. Mammalian GEFs that are Kiss-Toth. By bringing Sos to the membrane, the signalling cascade can be initiated, which will enable development at the non-permissive temperature of 37 ◦C, by stimulating the Ras transition. Consequently, a bait (X) that has been fused with Sos can detect membrane proteins (Y, prey) that interact with the bait. The system's output is dependent on Saccharomyces cerevisiae's capacity to grow and survive when the temperature is raised from 25 ◦C to 37 ◦C. Tavernier et al. (2002) detailed a method for studying protein-protein interactions in human cells that is conceptually comparable to the yeast two-hybrid system. Their research led to the development of the MAPPIT system, which allowed for the genetic modification of the JAK/STAT pathway at the level of the type I cytokine receptor [10]. The leptin receptor that they utilised had a Y1138F mutation, which prevents the receptor from recruiting the STAT transcription factor. As a result, the receptor that is attached to the ligand is unable to send out a signal. A bait and prey vector was introduced to the system to compensate for this signalling deficit. The bait is engineered onto the C-terminal of the Leptin receptor, and the prey is fused with the C-terminal section of gp130, which has four functioning STAT3 recruitment sites. The C-terminal region of gp130 can recruit STAT3 once prey binds to the bait. Then STAT3 is activated after being phosphorylated. After becoming active, the transcription factor can enter the nucleus and start expressing a reporter gene, like luciferase. Building a cDNA library expressing proteins in fusion with the GAL4 AD and the test protein (bait) coupled to the BD domain allows one to discover new proteins that interact with a target protein. Nevertheless, automation is frequently necessary for large-scale screenings that aim to systematically find interacting proteins against many baits, or even full signalling cascades. Below, we will outline the two primary methods that have been created to facilitate the deployment of robotic platforms. The library

screening method and the matrix (or array) approach are two examples of these [11, 12]. Matrix yeast screening involves mating a predetermined group of yeast clones that have had cDNA inserts inserted into them as fusions with the BD or AD. Research into protein-protein interactions in Saccharomyces cerevisiae and the cell cycle regulators of the fruit fly was carried out using this method. Despite the fact that Ito et al. and Uetz et al.'s research did not provide an entirely overlapping dataset, they did both make significant contributions to building the yeast Protein Interaction Map, which improves our knowledge of gene function in a single cell system. The second method involves screening one library against another, for instance to understand protein interaction maps in bacteria like Helicobacter pylori, by comparing random libraries or libraries of pooled open reading frames. It takes a lot of work, but researchers are still trying to list every potential protein-protein interaction in worms and yeast genomes. It is more difficult, time-consuming, and demanding to comprehensively finish the protein-protein interaction map in more sophisticated creatures (like mammals) with more complicated genomes and a variety of specialised cell types. Finding new proteins that interact inside a certain physiological signalling pathway or activity is, thus, a more practical approach to characterising a network of protein-protein interactions. For instance, knowing all the proteins that play a role in spliceosome function has helped us understand pre-mRNA splicing and has also given us a good starting point for investigating other cellular machinery. To sum up, the two-hybrid system has been around for a long time and is a vital tool for cell biologists. It can pick up on brief and weak interactions because of its great sensitivity. A single methodology may be used to detect and characterise protein-protein interactions, and the experimental setup is simple enough that anyone can do it. Without any prior knowledge, it can be used to find new proteins interacting with the bait protein [13]. By doing the experiment in an in vivo system (such as yeast or E. coli), the screen has a better chance of detecting real interacting proteins since the target proteins are more likely to fold correctly. Though there are certain restrictions with the two-hybrid system as well. The most significant technical issues can be either false-positives or false-negatives, both of which can happen. The two-hybrid technique is prone to false-positive detection when proteins interact with the promoter sequence or the DNA upstream of the reporter gene, among other potential sources of false-positive results. To illustrate the point, RNA polymerase II was the target of the initial two-hybrid system's design. Consequently, finding new proteins that interact with the activators of RNA polymerase II (as bait) is a challenging task. Some fusion proteins with BD or AD may not localise in the yeast nucleus, fold incorrectly, lack functionality when expressed as a fusion protein, be harmful to the host, fail to undergo the proper posttranslational modification, or be underrepresented in the library, which can lead to false-negativity. In order to validate and confirm the hits found by the two-hybrid system, it is recommended to utilise other approaches simultaneously, such as procedures based on biological function.

Electrophoresis in Two Dimensions:

The two-dimensional gel electrophoresis was the gold standard for getting a genome's overall translation profile in the early days of proteomics. The great resolution is one of the main characteristics of the method. As is typical with regular SDS-PAGE, a protein mixture is initially separated in one direction by isoelectric focusing and then in the other way by molecular weight. 2D electrophoresis is a powerful tool in proteomics, despite being a comparatively established technology. When a protein of interest is immunoprecipitated and then subjected to 2D electrophoresis, its interaction partners can be identified. It is common practice to combine two-dimensional gel electrophoresis with affinity chromatography and mass spectrometry to better identify the different gel spots .

Gas Chromatography:

Mass spectrometry, sometimes called mass spectroscopy, mass-spec, or MS, is a type of analytical technique that generates a mass spectrum—a unique property of a physical sample—to determine the mass-to-charge ratio of ions. Mass spectrometry has lately exploded in popularity as a tool for protein analysis. Ionisation, mass separation, and detection are the three main parts of a mass spectrometer. Only gaseous molecules with charges may be separated by it, and even then, only molecules with positive or negative charges can be separated simultaneously. The two main techniques for ionising whole proteins are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionisation (ESI). Under normal air pressure, ESI can convert liquid samples to gaseous state. Co-crystallization of an acidified matrix (a tiny molecule) with samples of interest (peptides and protein complexes) is a MALDI ionisation process. Electrospray allows for the preservation of non-covalent connections between molecules due to its mild ionisation. This means that huge assemblies like viral capsids and heterogeneous, asymmetric ribosomes can be ionised and kept intact, in addition to simple dimers. Matrix absorption occurs in the UV region [13, 14], and the absorbed energy is thermally dissipated. Several charged proteins and peptides of interest end up in the gas phase as a consequence of this dissipation. Since mass analysis of proteolytic peptides allows for the use of less complex instrument designs, leading to more affordable setups, it has become a significantly more prevalent approach for protein characteriszation. The digestion of entire proteins into smaller peptide fragments also makes sample preparation easier. When it comes to peptide mass analysis, the quadruple ion trap is the tool of choice. This application also makes use of MALDI time-of-flight instrumentation and multiple stage quadruple-time-of-flight instruments. Combining MS with affinity purification, high-performance liquid chromatography (HPLC), or two-dimensional SDS-PAGE allows researchers to investigate protein-protein interactions or identify proteins within complexes. In order to research protein-protein interactions, one common way to utilise affinity purification is to attach an epitope, like FLAG (sequence DYKDDDDK) [15], to the N-terminal end of a gene that is of interest. This will help find new proteins that interact with the bait protein, which is the result of this gene. After transfecting a vector with the FLAG-tagged gene into a mammalian cell line (like HEK293), the cell extract is passed through a column that has an antibody covalently linked to it (anti-FLAG in this case). This allows for the targeted recovery of proteins that interact with the bait protein. Two main methods employ mass spectrometry (MS) to identify proteins [15]: peptide mass fingerprinting and tandem MS/MS. To conduct peptide mass fingerprinting, the masses of proteolytic peptides are fed into a database that stores the projected masses that would result from an in silico digestion of a computer-generated list of known proteins by the same enzyme, often trypsin. This method sorts the database entries for proteins by the amount of peptide masses that correspond to their expected trypsin digestion pattern. Using a database of anticipated masses for one of several given peptide sequences, MS/MS is rapidly replacing other experimental methods as the go-to way to identify proteins. Further, MS/MS has the ability to identify complexes with large molecular weights, making it a useful tool for investigating protein-protein interactions. Using MS/MS, we can learn about the architecture, subunit stoichiometry, and composition of multiprotein complexes, such as mega-Dalton particles. Furthermore, the mass/charge ratio of an unmodified peptide is different from that of a peptide that has undergone post-translational modification, such as phosphorylation or covalent tagging with a polysaccharide. This variation can be identified by mass spectrometry (MS/MS) and subsequent analysis with software developed to find and identify peptide modification sites. Because of its usefulness in protein identification and characteriszation of post-translational changes, MS/MS has thus become an integral part of proteomics. Assemblies of ten or

more proteins mediate numerous crucial cellular biological processes .

Nuclear magnetic resonance (NMR) and X-ray crystallography are two of the more conventional structural biology platforms that work well with the MS method. Using MS/MS, one may not only determine the molecular make-up of complexes involving many proteins, but also locate the substrates or ligands that are bound to very diverse and complicated assemblies. Unambiguous protein identification and precise peptide and protein mass measurements are MS's strongest points when compared to other methods for investigating protein-protein interactions [16]. Using MS/MS as an example, we find new proteins in signalling networks. Members of the NF-κB transcription factor family, including p50/NF-κB1 and p65/RelA, which control the expression of numerous inflammatory genes and the 10 H inhibitor family, had been identified by 1997. Qian • E. The transcription factors known as IκB are Kiss-Toth. The activation of NF-κB, which allows the transport of p65/RelA into the nucleus, is initiated by the phosphorylation and degradation of IκB, which was previously reported. Despite this, the IκB kinases were still not identified and studied. Mercurio et al. (1997) used MS, chromatography, and bioinformatics-searching of EST databases to identify IKK-1 and IKK-2 as the IκB kinases. An IKK signalsome, a protein complex including TNF-α inducible IκB kinase activity, was segregated by gel-filtration chromatography from a TNF-α stimulated HeLa whole cell extract. Using high mass precision MALDI peptide mass mapping, the active fraction was separated by SDS-PAGE after chromatography. Two bands, measuring 85- and 87-kD, were excised and digested by trypsin. According to the peptide sequence study, the kinases that are relevant for phosphorylating IκB are IKK-1 and IKK-2.

Engineering using Proteins:

The experimental technique of modifying proteins to optimise their specific features is known as protein engineering. Some examples of these properties are protein-protein binding affinity and catalytic activity in "non-physiological" environments. Such endeavours necessitate knowledge from a wide range of fields, including genetics, protein biology, bioinformatics, mathematics, and in silico protein design. When it comes to protein engineering, there are two primary approaches. One method is rational design, which relies on in-depth understanding of the target protein's structure and function to achieve the necessary modifications. In order to impart specific structural changes to the target protein, site-directed mutagenesis techniques are utilized [17]. A new mutant protein can be more easily designed with the aid of computational protein design techniques. Nonetheless, this approach might not be the best choice when dealing with newly discovered proteins or when a high-resolution model of the protein's structure is unavailable. Liu et al. (2007) recently described an example of rational design in action. The authors of this work used computer algorithm modelling to explore the possibility of computationally redesigning protein-protein interactions by moving functional epitopes from one protein to another. Their goal in redesigning the protein-protein interface was to make it bind to the erythropoietin receptor in a cell-based experiment. To achieve this, they produced a mutant rat pleckstrin homology domain of phospholipase C-δ1. The non-natural protein-protein interface was successfully examined for binding affinity and biological functions (using luciferase as a reporter readout). Their research proved that computational redesign can be a powerful tool for creating protein-protein interaction pairings that do not occur naturally. "Directed evolution" describes the second tactic. This method involves creating a pool of protein mutants using random mutagenesis and then selecting for those mutants that meet the necessary criteria. It is possible to apply further rounds of selection to produce more mutants if required. For research involving protein-protein interactions, this approach is suitable for high-throughput experiments [18]. Using this method, one can create new enzymes

and high-affinity antibodies (for use in enzyme engineering and biocatalysis) from protein and peptide libraries.

To find inhibitors of specific protein-protein interactions, other methods that are commonly called protein engineering include creating small molecule leads, cross-linked interfacial peptides, and α-helix mimetics. An crucial death domain of the Bcl-2 members, the BH3 peptide, was created by Walensky et al. via a chemical technique they called hydrocarbon stapling. The compounds known as "stabilised α-helix BH3 peptides" were found to be helical, resistant to proteases, and able to pass through cells. They bound to pockets of multidomain Bcl-2 with a higher affinity. A potential tool for exploring modulations of protein-protein interactions could be these chemically modified peptides. Likewise, in order to make the α-peptide more stable and resistant to proteolysis, Kritzer et al. chemically replaced a residue on the p53 trans-activation domain. Following the alteration, the peptide had the potential to trigger cell death. The nature of protein-protein interactions can be better understood with the use of these methods.

Table 1. Physiochemical experimental techniques identify the physiochemical interactions between proteins (Protein-Protein Interaction Techniques).

No.	Method	Information
1.	**Yeast two-hybrid**	Analyzing protein–protein interactions using the two-hybrid system
2.	**Tandem affinity purification**	The tandem affinity purification (TAP) method: a general procedure of protein complex purification
3.	**Mass spectrometry**	Systematic identification of protein complexes in *S. cerevisiae* by mass spectrometry
4.	**Immunoprecipitation**	Complex formation between the UL16 and UL21 tegument proteins of pseudorabies virus
5.	**Pulldown assay**	Protein interactions among the vaccinia virus late transcription factors
6.	**Phage display**	Protein–protein interactions in hematology and phage display
7.	**Protein chips**	Global analysis of protein activities using proteome chips

However, only few detection methods for protein-protein interactions can be easily adapted for a high-throughput strategy. These include : ((Yeast two-hybrid and Affinity purification/mass spectrometry))

1. **Yeast two-hybrid :**

- The Y2H technique allows detection of interacting proteins in **living yeast cells**. Interaction between two proteins, called **bait** and **prey**, activates reporter genes that enable growth on specific media or a color reaction.
- Y2H can be easily automated for highthroughput studies of protein interactions on a genome-wide scale, as shown for viruses like **bacteriophage T7**, ***Saccharomyces cerevisiae***, ***Drosophila melanogaster***, ***Caenorhabditis elegans*** and **humans**.

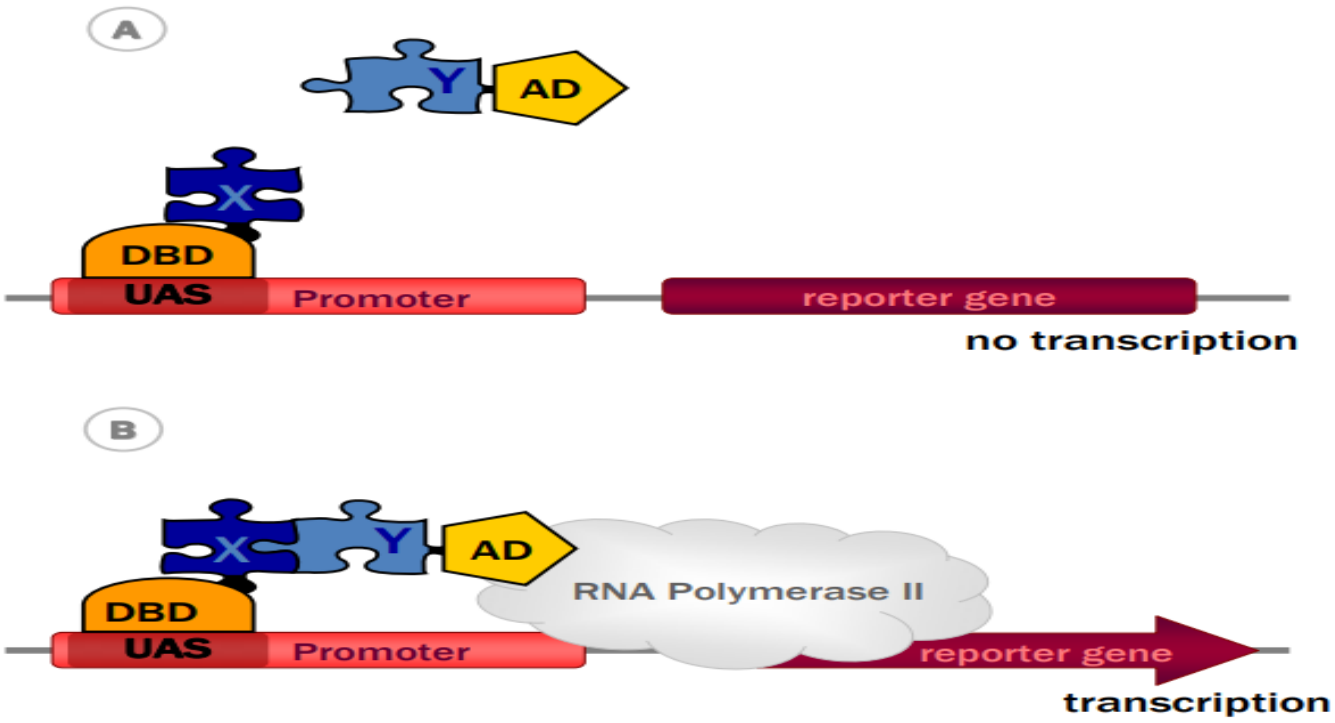
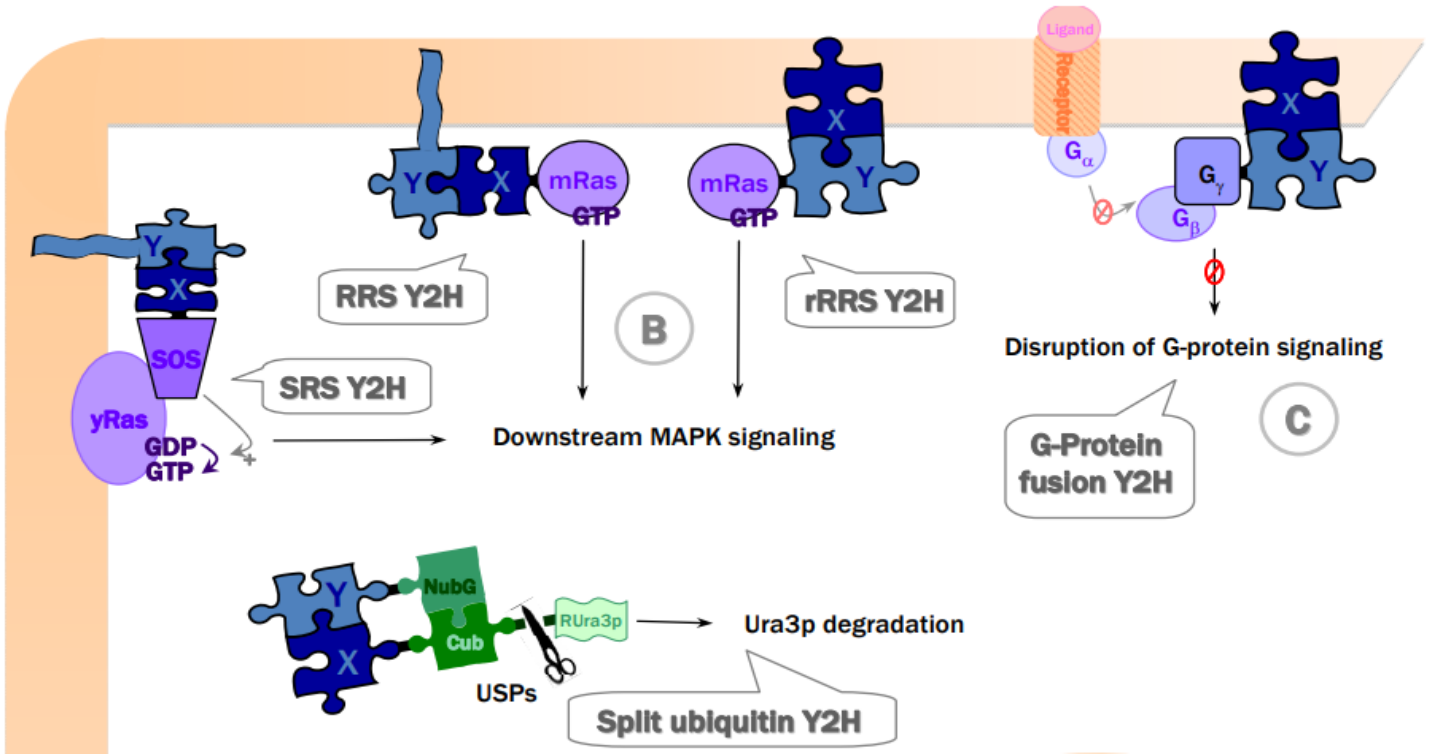


Figure 1. The classical yeast two-hybrid system.

(A) The protein of interest **X**, is fused to the DNA binding domain (**DBD**), a construct called **bait**. The potential interacting protein Y is fused to the activation domain (**AD**) and is called **prey**.

(B) The bait, i.e. the **DBD-X fusion protein**, binds the upstream activator sequence (UAS) of the promoter. **The interaction of bait with prey, i.e. the AD-Y** fusion protein, recruits the AD and thus reconstitutes a functional transcription factor, leading to further recruitment of **RNA polymerase II** and subsequent **transcription of a reporter gene.**



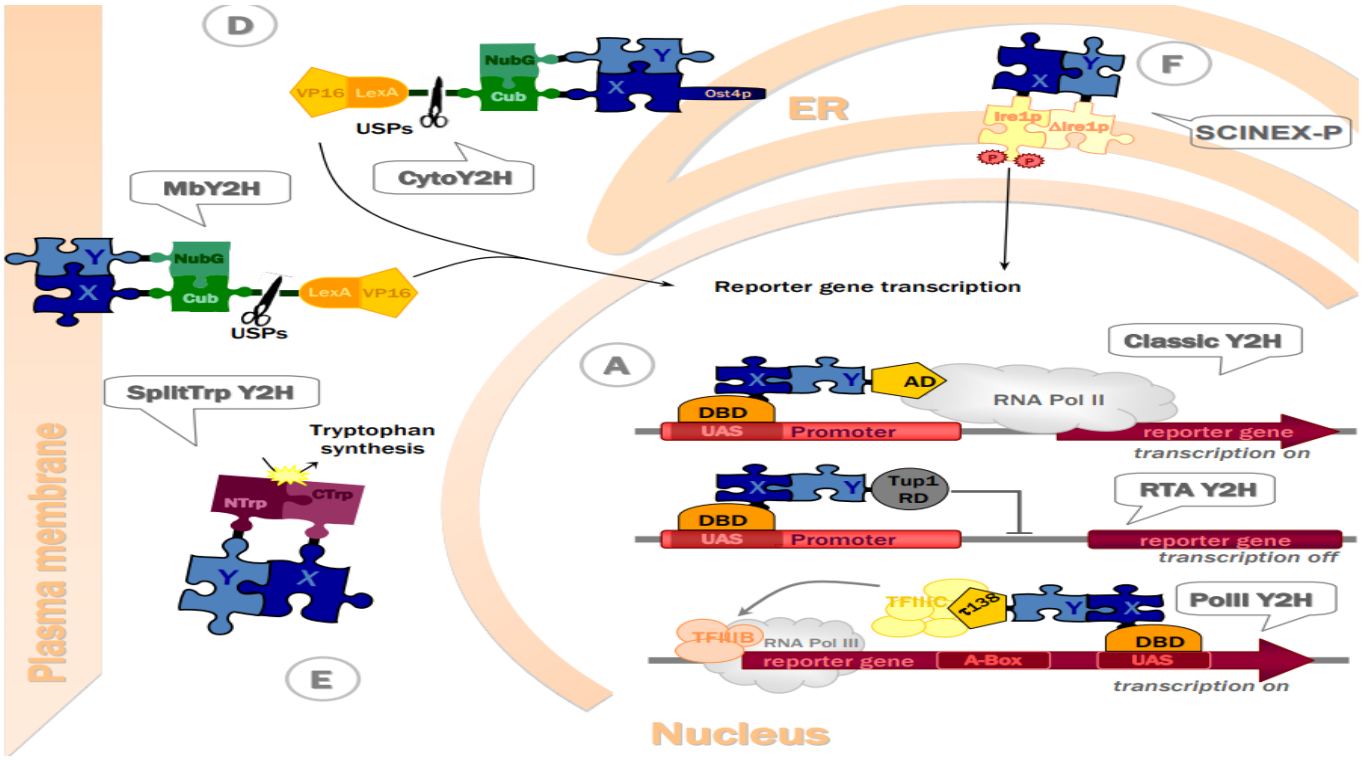


Figure 2. Yeast two-hybrid systems, their subcellular location within a yeast cell, and their operating mode (represented at the moment of bait-prey interaction).

Protein X (dark blue puzzle piece, part of bait construct) and **protein Y** (light blue puzzle piece, part of prey construct) directly interact (fitting puzzle pieces), thus inducing reconstitution of split proteins (puzzle pieces of different colors in **A, D, E**), membrane recruitment **(B, C)** or protein dimerization **(F).** Protein fusions in bait or prey constructs are shown as solid black lines between puzzle pieces. Bait-prey interaction activates further downstream events (arrows) that **directly (A)** or **indirectly (B, C, D, F)** lead to transcriptional activation, or are independent of transcriptional activation **(D, E),** finally yielding screenable readouts like growth on specific media or color reactions.

(A) Nuclear Y2H systems all require protein recruitment and bait-prey interaction at nuclear DNA. The classic Y2H and RTA Y2H both engage RNA polymerase II (RNA Pol II) transcription either by its activation or its inhibition. By contrast, the Pol III Y2H, involves RNA polymerase III (RNA Pol III) transcription.

(B) Ras signalling based Y2H at the plasma membrane. The SRS Y2H, RRS Y2H, and rRRS Y2H are all based on protein recruitment to the plasma membrane via bait-prey interaction and subsequent activation of MAPK downstream signalling. While in the SRS and RRS Y2H the prey constructs harboring protein Y are anchored at the membrane via myristoylation to analyze interactions with cytosolic bait constructs harboring protein X, the rRRS is used to analyze interactions between soluble preys containing protein Y and partner X being a membrane protein.

(C) G-protein signalling-based Y2H at the plasma membrane. In the G-protein fusion Y2H, bait X is a membrane or membrane-associated protein whose interaction with the prey construct disrupts protein G downstream signalling.

(D) Splitubiquitin based Y2H systems involve reconstitution of ubiquitin from two domains upon bait-prey interaction. Their subcellular localization depends on the nature of interacting proteins X or Y, and on the reporter proteins used. The Split ubiquitin Y2H uses non-transcriptional reporting of protein interactions in the cytosol, but can also be used for membrane proteins (not shown).

The MbY2H is used for interaction analysis with membrane baits and thus occurs at the membrane location of protein X, e.g. the plasma membrane. The CytoY2H is used for membrane anchored cytosolic baits and occurs close to the ER membrane

(E) Split-protein sensor Y2H. The Split-Trp Y2H is used to assay cytosolic bait-prey interactions based on reconstitution of an enzyme in tryptophan synthesis, allowing for non-transcriptional reporting.

(F) ER Y2H system. The SCINEX-P Y2H allows baitprey interaction analysis in the reducing environment of the ER, based on protein dimerization in unfolded protein signalling. ER, endoplasmic reticulum.

2. Affinity purification/mass spectrometry

- The value of mass spectrometry for high-throughput screening of protein interactions has been recognized only more recently.
- **Mass spectrometry analytical technique is based on** the determination of the mass-to-charge ratio of ionized molecules.
- These include Nobel prize crowned methods for ionization like **electrospray ionization**, **generating ions from macromolecules in liquid medium without their fragmentation**, **soft laser desorption (SLD)** or **matrix-assisted laser desorption/ionization (MALDI)**, using a laser beam for ionization of macromolecules without breaking chemical bonds.
- **Mass spectrometry is now routinely applied to identify proteolytic fragments of proteins or even entire proteins and protein complexes.** Coupled to classic biochemical methods like affinity purification or chemical cross-linking, MS has become also a powerful tool for large-scale interactome research, mainly in form of affinity purification-MS (AP/MS).

Comparison of Y2H- and MS-based methods:

Mass spectrometry is less accessible than Y2H due to the expensive large equipment needed. Thus, a large amount of the data so far generated from protein interaction studies have come from Y2H screening. For example, more than 5,600 protein interactions have been so far reported for yeast and about 6,000 for humans, establishing extensive protein interaction networks.

Pull-Down Experiments in the Lab

Many techniques rely on the same basic idea: co-precipitation of protein X and its associated proteins. Consequently, as will be shown later on, in vitro pull down is employed in a number of designs. By inserting many genes from a "gene bank/library" into bacteriophages, a test called phage display can be performed to screen for protein-protein interactions. Placing a target protein (X) on a plastic dish is the fundamental step in phage display. A bacteriophage's coat-protein and a gene bank from an organism's genome are fused in a library, allowing the two to coexist on the viral particle's surface. After that, the coated dish is supplemented with the phage display library. No particles of the phage showing proteins that interact with protein X can be washed away; only those remaining attached to the dish will be retained. Along with the encoding gene, the phage particles carry the genetic material. This establishes a causal relationship between genotype and phenotype. By sequencing the encapsulating DNA, the sequences of the interacting peptides can be found. Two methods exist for presenting library members, such as peptides, to the M13 phage coat proteins. A polyvalent display is achieved by first displaying on the main coat protein (protein-8). Optimal engineering allows for one display in each phage particle when displaying on the minor coat protein (protein-3), the second option. Deshayes et al. (2002) provided an example of how to use phage display to uncover high affinity ligands or peptides to a specific protein. They zeroed in on insulin-like growth factor, a hormone with 70 residues, as their target to research in order to identify its different epitopes. They proved that from a huge peptide library, they were able to extract identifiable peptide motifs that bound to the insulin-like growth factor receptor using their improved phage display approach.

the Protein Fragment Complementation Assay (FRET (

Energy transfer between two chromophores, where the emission energy of one molecule (the donor) overlaps with the excitation energy of another (the acceptor), is the basis of fluorescence resonance energy transfer (FRET), a popular tool for studying protein-protein interactions. A portion of the excited energy is transferred to the second molecule when the donor chromophore is stimulated at its unique fluorescence excitation wavelength. Hence, in order to implement this concept in studies of protein-protein interactions, the donor molecule tags the target protein (X), while the acceptor molecule tags either a test protein or a cDNA library. Fluorescent microscopy or flow cytometry can detect the fluorescence released from the acceptor when the proteins under research interact, which brings the donor and acceptor into close contact (1-10 nm). Interactions between proteins, between proteins and DNA, and between proteins and changes in their shape can all be studied using FRET. For research on intracellular proteins in particular, green fluorescent protein (GFP) has served as a metric for cellular physiology. Because it allows the real-time measurement of protein-protein interactions using live cells, the cyan fluorescent protein (CFP)-yellow fluorescent protein (YFP) pair is the most used FRET pair for biological applications [19]. A unique variation of the green fluorescent protein (GFP) is shared by both CFP and YFP. Another common pair of proteins is BFP and eGFP, which stands for enhanced green fluorescent protein; both proteins are variations of GFP. Nevertheless, due to its weak fluorescence, the BFP is not well-suited for uses outside of fluorescent microscopy and flow cytometry. Results from Advanced Technologies for Protein-Protein Interaction Studies 13 may be heavily impacted by photo-bleaching, or the need to directly excite the acceptor, due to the fact that FRET relies on external illumination to start the fluorescence transfer. An alternate method, bioluminescence resonance energy transfer (BRET), has been devised to overcome some of these restrictions. BRET works by first excitation of a GFP protein variant—for instance—by a bioluminescent luciferase, which acts as an energy giver. YFP (using it as a source of energy). Being more suitable for imaging small animals, BRET produces energy by chemiluminescence, making it more sensitive in live experimental objects than FRET systems (De and Gambhir 2005). A relatively new approach to investigating protein-protein interactions is the protein fragment complementation assay (PCA). PCA makes use of two pieces extracted from a single fluorescent protein (e.g., GFP, YFP) on the theory that certain transcription factors or fluorescent proteins are modular and can reassert their activity when their two domains are near to each other. A protein of interest is fused to one half of the fluorescent protein, which is typically the YFP. A functioning fluorophore can be formed from the two YFP fragments if the proteins bind to one another. In addition to discovering new protein-protein interactions, principal component analysis (PCA) is helpful for studying how the introduction of other agents affects existing connections. Our group and others have used principal component analysis (PCA) to investigate protein roles and visualise interprotein interactions. We have used principal component analysis (PCA) to investigate MKK-Tribble interactions in primary smooth muscle cells and epithelial cell lines. The use of principal component analysis (PCA) in a cDNA library screen to map PKB signalling networks was reported by Remy et al .

Novel components of the signalling systems activated by PKB were characterised as a result of these investigations. Although FRET and PCA technologies necessitate comparable equipment, they are complementary due to their distinct advantages and disadvantages. FRET enables tests of protein-protein interactions in living cells with high spatial resolution. For fluorescence resonance energy transfer (FRET) to take place, the distance between the donor and acceptor fluorophores must be smaller than 10 nm. Direct protein-protein interaction also takes place on a comparable spatial scale. Having

said that, FRET does not respond very well to certain conditions and has a very limited dynamic range. For the two fluorophores to work together, their fluorescence intensities must be comparable. The two components of FRET must be expressed in the same cells at optimal quantities, which makes large-scale investigations challenging to do without extensive automation. A ratio of 10:1 to 1:10 for donor/acceptor expression is required. It could be challenging, for instance, to ensure that every single cDNA in a cDNA library screen is expressed to the bait fusion protein at an optimal quantity. In the absence of alignment or proximity of less than 10 Angstroms between the two fluorophores, the FRET signals may go unnoticed. This may lead to 14 H that are not real. Qian • E. Negative for Kiss-Toth interactions, even when target proteins bind but fluorophores are misaligned in complex. Additionally, the two fluorophores should not interact with one another in order to avoid false-positive readings. Signal cross-talk between the donor and acceptor fluorophores occurs during FRET image acquisition due to the overlapping of the two fluorophores' emission spectra. According to Piston and Kremers (2007), optimising the genetic alteration of fluorophores and operating the imaging acquisition can be achieved by doing things like creating GFP variants that are as good as they can be and fine-tuning the fluorescent microscope. Whereas optimising protein expression levels is not necessary for the two segments to form an active 3D structure using PCA, scaling up the process is simpler. And because PCA works by observing how fluorescent proteins fold, the dynamic range of a fluorescent signal is also very high. Since a third protein can bridge the gap and bring the two proteins into close enough proximity to be detected, a PCA signal does not differentiate between direct and indirect protein-protein interactions. Nevertheless, research into the building of protein interactomes finds the method particularly helpful. So, to confirm and provide more details on the protein-protein interactions found by PCA, further approaches like co-immuno-precipitation would be required.

RNAi Down-regulation

A strategy for reducing gene expression known as RNA interference (RNAi) has just recently been discovered, but it has evolved to function as a genome protector for the vast majority of eukaryotic creatures. Evidence of RNA interference (RNAi) was initially found in 1986 in transgenic plants, when antisense RNA that had been transfected prevented the transcription of homologous messenger RNA. In the early 1990s, researchers saw that plants and fungus could downregulate RNA expression from both external sources and within. Fire et al. (1998) documented the nematode C. as the first organism in which RNA inhibition was observed in animal cells. they noticed in elegans and came up with the phrase "RNAi" from there. Following the introduction of lengthy double-stranded RNA (dsRNA) into C. Selective degradation of homologous mRNA occurred in C. elegans. Over the last decade, research into the molecular mechanism of RNA interference (RNAi) has exploded in popularity among biologists. Evidence suggests that short double-stranded RNAs (dsRNAs) ranging in length from 21 to 25 nucleotides are the effector molecules responsible for mRNA degradation; thus, the name "small interfering RNA" (siRNA). DICER cleavage or exogenous synthesis are two methods by which lengthy dsRNAs can be converted into siRNAs. Aside from 15-long dsRNA, synthesised siRNA can also be introduced into cells to trigger RNA degradation (Advanced Technologies for Protein-Protein Interaction Studies). The RNA-induced silencing complex (RISC) can identify both endogenous and exogenous siRNAs. The siRNA stops translation of a specific mRNA after binding to RISC by targeting it and then triggering its cleavage. Nature provides a variety of small RNA species, such as siRNAs, miRNAs, and rasiRNAs, which are connected with repeats. Small RNAs develop via distinct routes and each has its own unique set of targets for degradation or suppression. In some creatures, like the worm-like nematode C. Although the generic inhibitory mechanisms are comparable in worms, fruit flies Drosophila melanogaster, and

flowering plants Arabidopsis thaliana, the RISC complex and enzyme dicer are unique to each of these species. Since there have been recent, extensive reviews of RNAi's biological mechanism, we shall limit our attention to its use in investigating protein-protein interactions. Gene function has been studied in various animal model organisms, including C. elegans, using genome-wide synthetic RNAi libraries. species, including fruit flies (Drosophila), worms, and hexaploid wheat (Triticum aestivum) and Arabidopsis thaliana (plants). A complete library of about 500 base-pair dsRNAs that covers over 90% of the Drosophila mRNA could be utilised for high-throughput screening to find genes related to cell viability, as shown by Boutros et al. Of the 438 dsRNAs found in this study, identifying important genes, 80% did not have mutant alleles. Three methods exist for successfully introducing double-stranded RNA into worm C. in order to determine the roles [20] of genes in elegans through injection, soaking, and feeding. In order to feed C., multiple trials were conducted. of the genus E. Escherichia coli producing double-stranded RNA for the purpose of identifying genes of interest with loss-of-function. Using this strategy, Kamath et al. generated a library of 17,000 siRNA constructs, or 86% of the C. elegans genome, yielding the discovery of over 1700 mutant phenotypes, with a significant portion of them being unique. The worldwide reduction of gene expression caused by the dsRNA-induced activation of the interferon response later rendered lengthy dsRNAs useless in mammalian cells. In contrast, small interfering RNAs (siRNAs) (21-25 nt) are able to screen libraries in mammalian systems without often stimulating the interferon response. Robots and automated data analysis have made possible the development of new methods for building siRNA libraries for use in high-throughput screening in conjunction with microarray/microwell platforms. Not long ago, Zhao and Ding documented an instance of screening cDNA libraries employing siRNAs (2007). To find naturally occurring regulators of osteogenic specificity, the researchers used a synthetic siRNA library that targeted five thousand human genes. The research found 53 potential suppressors; 12 of them were later proven to play critical roles in preventing human mesenchymal stem cells from undergoing osteogenic differentiation. While the exact molecular mechanism of RNAi in human cells remains a mystery, experimental biologists have made use of the phenomenon to investigate gene functions in both in vitro and in vivo model animals.

Screening for cDNA Libraries

Gene discovery by cDNA library screening is an all-encompassing approach to genomics and proteomics that makes use of several approaches, some of which have already been discussed. The goal of this work is to map intracellular signal processing pathways and discover new signalling proteins using cDNA library screening technologies that we created in our lab. Genes are identified using our cDNA library screening method according to their function within a certain signalling network. It has been shown that cloning the gene expression of mammals is an effective method for studying the interactions of chemicals within cells. Aruffo and Seed's (1987) methods were used to clone IL1R almost twenty years ago, however a modified technique called single cell autoradiography following radioligand binding was used instead. Our modification to this method for signalling pathway mapping involved combining a cDNA expression pool with a transcription reporter using co-transfection. This process was dubbed "Transcription Expression Cloning" by us. It takes advantage of the idea that a sensitive reporter system can detect the downstream response when most signalling components are overexpressed, just like extracellular agonists. Utilising the IL8 promoter—which includes NF-κB, c/EBP, and AP-1 sites—to drive the reporter enhanced green fluorescent protein (EGFP) or luciferase, we utilised this technique to discover new components in TIR receptor-induced signalling cascades. Although luciferase was utilised as a reporter (since it was more easily automated), proof-of-concept experiments demonstrated that EGFP as

a reporter can be utilised to map pro-inflammatory signalling pathways using confocal analysis. The HeLa cells were co-transfected with three different plasmids: (1) an IL8 promoter-driven firefly luciferase reporter (IL8-Luc), a Renilla luciferase reporter driven by the HSV Thymidine kinase promoter (pTKrLuc), and a set of cDNA expression library clones (oligo-dT primed, non-directionally cloned, from a human peripheral blood mononuclear cells) driven by the CMV promoter. To find the bioactive cDNAs, DNA pools were tested and the positive pools were further split down. In genome-wide screens, this method helped us and others find several new components involved in various signalling cascades.

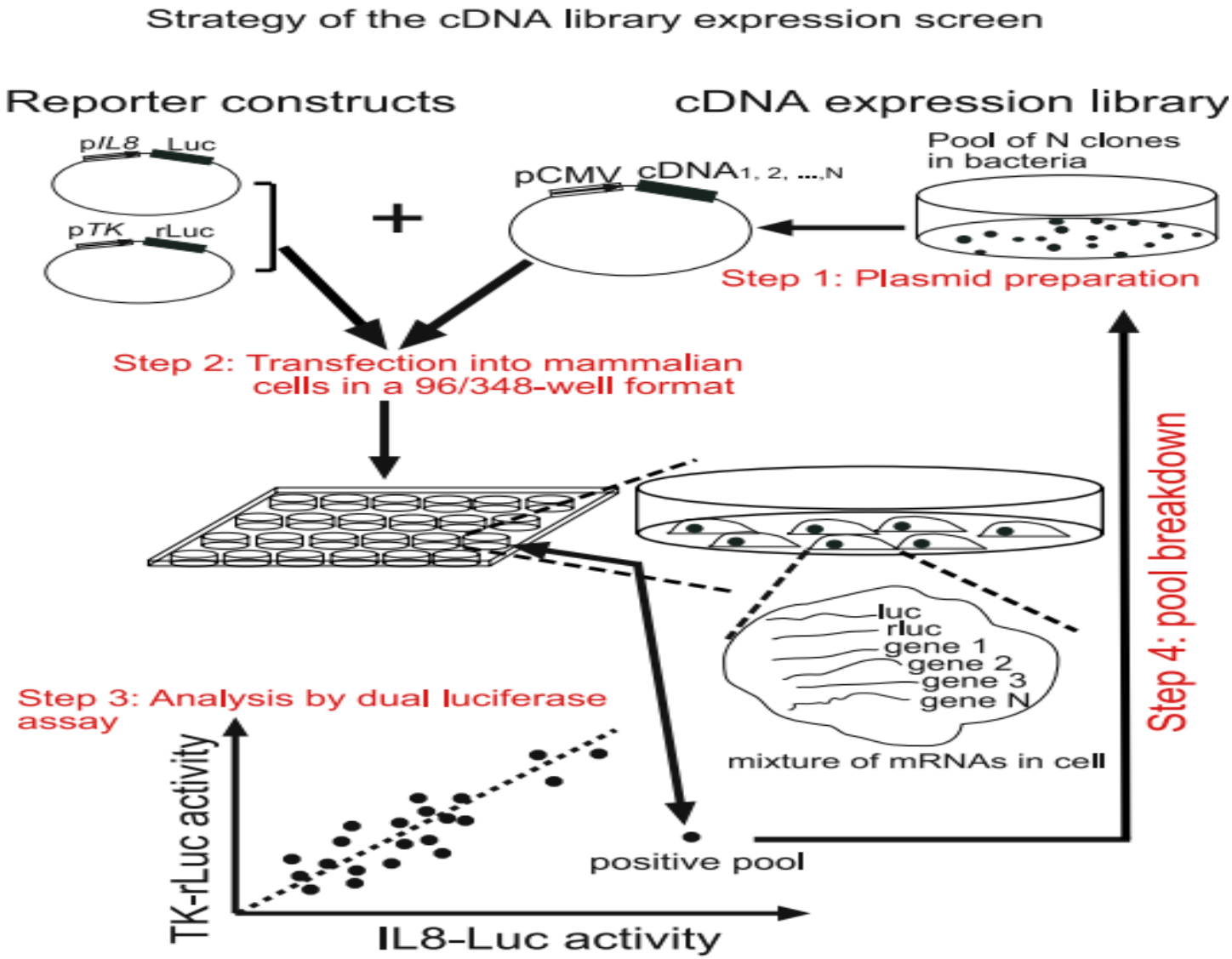


Figure 3. Plan for our earlier cDNA library screening. A cDNA library consisting of approximately 3,106 clones was divided into N pools. The size of each pool was decided by finding the minimum amount of cDNA needed for detection in step 3. The plasmid pools were introduced into mammalian cells (such as HeLa) in 96-well plates together with reporters (pIL8-Luc and pTK-rLuc). A dual luciferase assay (Promega) was used to analyse the reporter activity after 24 hours, and positive pools were detected. In order to find the bioactive cDNA clone, the positive pools were first subdivided into smaller subpools, and then into individual cDNA clones. Then, processes 1-3 were repeated.

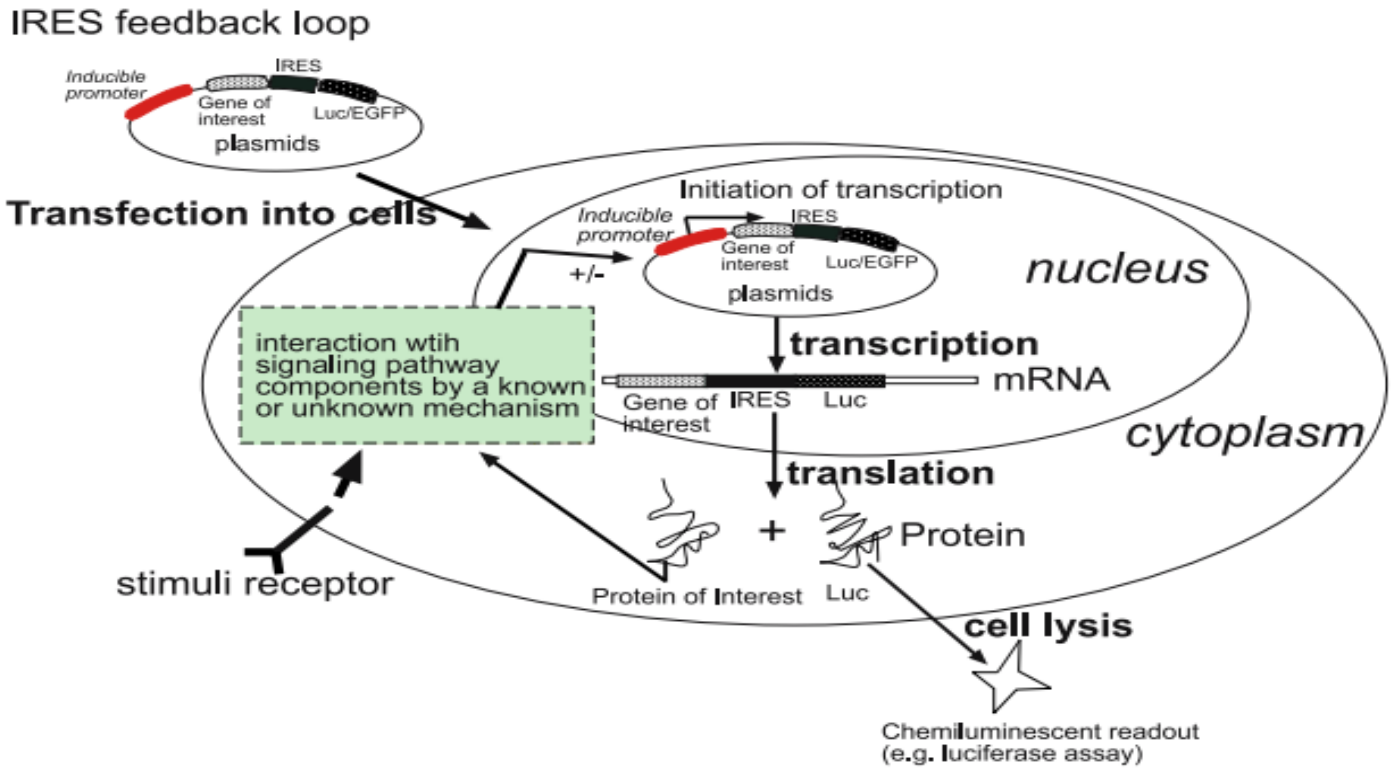


Figure 4. Loop for IRES feedback. For the purpose of screening cDNA libraries, this technique makes use of a modified IRES vector. An IL-8 promoter, which could be activated by inflammatory signals, regulated the expression of cDNA clones and reporter genes (EGFP or luciferase). The reporter gene and the gene of interest were both expressed on the same cellular transcript, guaranteeing the simultaneous presence of both protein products. A positive (or negative) feedback loop forms when the protein, which is the product of the gene of interest in the IRES construct, is a regulator of a signalling network. When the IRES expression cassette is introduced into the cells, the reporter readout (e.g., luciferase assay) reflects the non-linear perturbation of the signalling system of interest.

Activated Protein Kinase networks. So, a high-capacity luminometer, a liquid handling robot, a high-throughput colony picker, and luciferase as a reporter make it possible to screen expression libraries containing 106 clones or more in a timescale that is realistic for typical research projects. Our ability to rapidly investigate protein interaction networks is greatly enhanced by these systems, and they can also quickly categorise genes according to their bioactivities. We used a strong viral CMV promoter to drive the components of a cDNA library in our expression cloning process, which has a limitation: if we express the component at a concentration significantly higher than its normal concentration, it could cause artificial biological reactions, which would lead to false-positives. The creation of both positive and negative feedback loops, however, is essential to signalling systems in order to accomplish non-linear regulation.

In controlling how the body reacts, these are crucial. Thus, in order to discover new components of the TIR signalling pathway, we have studied an improved version of the original technique that incorporates a positive feedback loop into our system. We have cloned the reporter (EGFP/Luciferase) into one of the expression cassettes and the potential regulator (Y) into the other using an IRES vector. This allows for the co-transcription of the two genes. Outlined in is the strategy. Using well-studied proteins like RelA,

TRAF6, MyD88, and IκBα, we have confirmed that this method may be employed to choose particular proteins within the TIR signalling pathway. One major improvement over earlier approaches is the use of an inducible promoter, rather than a robust viral promoter, to drive the transcription of both the reporter and the potential regulator from a cDNA library onto a single transcript. Consequently, there are less fake reactions since the test protein's expression is self-regulated and closer to the physiological concentration.

Conclusion:

Characterizing protein-protein interactions is essential for molecular understanding of gene function. In order for physiological interactions to occur or to be interrupted, it is often necessary to enzymatically alter complex components during the dynamic process of multi-protein complex formation. Deploying many experimental procedures is frequently required to obtain a thorough knowledge of these phenomena. We have compiled a list of the most popular approaches to determining gene function by analysing protein-protein interactions in this article. It is clear from the examples given above that automation and the creation of assays in a format that allows global examinations of the proteome constitute one of the primary development avenues for technologies. Although these systems are definitely advancing our knowledge of molecular and cell biology, research is out of reach for many people due to the exorbitant cost of the equipment needed to use many of the new technologies. Luckily, this is usually only a short-term restriction, as the price of instruments drops quickly when a new method is used more frequently. Furthermore, we think it's critical to create new tests that don't always necessitate a lot of money. The MAPPIT system, functional expression cloning, and tweaks to the classical Y2H are a few examples of such methods. These make it possible for the general scientific public to have access to state-of-the-art biological research, which guarantees the worldwide dissemination and accumulation of information.

References:

1. Kiss-Toth E, Qwarnstrom EE, Dower SK (2004b) Hunting for genes by functional screens. Cytokine Growth Factor Rev 15:97–102
2. Kiss-Toth E, Wyllie DH, Holland K, Marsden L, Jozsa V, Oxley KM, Polgar T, Qwarnstrom EE, Dower SK (2006) Functional mapping and identification of novel regulators for the Toll/Interleukin-1 signalling network by transcription expression cloning. Cell Signal 18:202–214
3. Knudsen CR, Jadidi M, Friis I, Mansilla F (2002) Application of the yeast two-hybrid system in molecular gerontology. Biogerontology 3:243–256
4. Kritzer JA, Lear JD, Hodsdon ME, Schepartz A (2004) Helical beta-peptide inhibitors of the p53-hDM2 interaction. J Am Chem Soc 126:9468–9469
5. Lettre G, Kritikou EA, Jaeggi M, Calixto A, Fraser AG, Kamath RS, Ahringer J, Hengartner MO (2004) Genome-wide RNAi identifies p53-dependent and -independent regulators of germ cell apoptosis in C. elegans. Cell Death Differ 11:1198–1203
6. Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP (2003) Vertebrate microRNA genes. Science 299:1540
7. Liu S, Liu S, Zhu X, Liang H, Cao A, Chang Z, Lai L (2007) Nonnatural protein–protein interaction-pair design by key residues grafting. Proc Natl Acad Sci USA 104:5330–5335
8. McGinnis K, Chandler V, Cone K, Kaeppler H, Kaeppler S, Kerschen A, Pikaard C, Richards E, Sidorenko L, Smith T et al (2005) Transgene-induced RNA interference as a tool for plant functional genomics. Methods Enzymol 392:1–24
9. Meng X, Wolfe SA (2006) Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. Nat Protoc 1:30–45

10. Mercurio F, Zhu H, Murray BW, Shevchenko A, Bennett BL, Li J, Young DB, Barbosa M, Mann M, Manning A, Rao A (1997) IKK-1 and IKK-2: cytokine-activated IkappaB kinases essential for NF-kappaB activation. Science 278:860–866
11. Michnick SW (2003) Protein fragment complementation strategies for biochemical network mapping. Curr Opin Biotechnol 14:610–617
12. Michnick SW (2004) Proteomics in living cells. Drug Discov Today 9:262–267 50. Muzio M, Ni J, Feng P, Dixit VM (1997) IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling. Science 278:1612–1615
13. Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell 2:279–289
14. Nelson RW, Dogruel D, Williams P (1994) Mass determination of human immunoglobulin IgM using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 8:627–631
15. O'Farrell PZ, Goodman HM, O'Farrell PH (1977) High resolution two-dimensional electrophoresis of basic as well as acidic proteins. Cell 12:1133–1141
16. Osborne MA, Dalton S, Kochan JP (1995) The yeast tribrid system–genetic detection of trans-phosphorylated ITAM-SH2-interactions. Biotechnology (NY) 13:1474–1478
17. Patil KR, Rocha I, Forster J, Nielsen J (2005) Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics 6:308
18. Piston DW, Kremers GJ (2007) Fluorescent protein FRET: the good, the bad and the ugly. Trends Biochem Sci 32:407–414
19. Pollok BA, Heim R (1999) Using GFP in FRET-based applications. Trends Cell Biol 9:57–60
20. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V et al (2001) The protein–protein interaction map of Helicobacter pylori. Nature 409:211–215