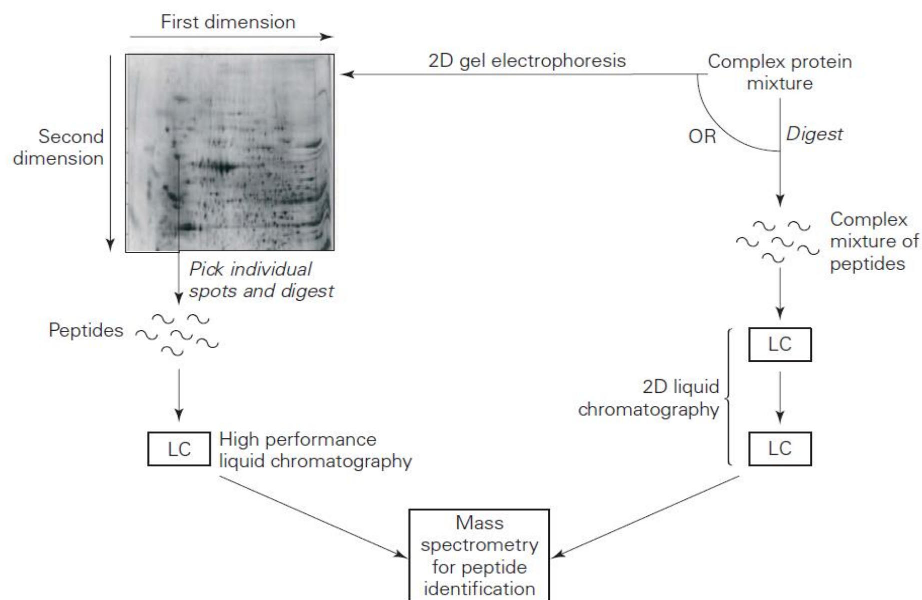
Proteomics

- Proteomics deals with the global determination of cell function at the level of the proteome. Proteome analysis is of crucial importance to our understanding of how cells function and of how function changes during disease and is of great interest to pharmaceutical companies in their quest for new drug targets.
- The first step is the extraction and separation of all the proteins in a cell or tissue. The most commonly used separation method is **two-dimensional gel electrophoresis**. The proteins in the extract are first separated in the first dimension according to charge in a narrow tube of polyacrylamide gel by **isoelectric focusing** .
- The gel is then rotated by 90° and the proteins electrophoresed in the second dimension into a slab of gel containing SDS, which further separates them by mass. The separated spots are then stained with a protein dye. A 2D map of protein spots is thus created which can contain many hundreds of visibly resolved protein species.
- Individual spots are then cut from the gel and each digested with a protease such as trypsin to produce a set of peptides characteristic of that protein. The peptides are then separated by high performance liquid chromatography in fine capillaries and individually introduced into an **ESI-quadrupole** or **MALDI-TOF MS** to determine their masses and so produce a **peptide mass fingerprint** of the parent protein.
- This is then compared with a database of predicted peptide masses constructed for all known proteins from the abundant DNA sequence information that is now available for many organisms. **Tandem mass spectrometry (MS/MS)**, where two mass spectrometers are coupled together, can give even more precise identification.
- Peptides from the digested protein are ionized by ESI and sprayed into the MS/MS, which resolves them, isolates them one at a time and then dissociates each into fragments whose masses are then determined.
- The detailed information available in this way can be used to determine the primary sequence of the peptides, which, when coupled with an accurate mass, gives unambiguous identification of the parent protein, or the sequence can be used in a BLAST search to identify homologous proteins. Many of these

procedures can be automated with high-throughput robots and so large numbers of proteins can be identified within the proteome of a given cell.

- MS procedures are particularly useful for identifying post-translational protein modifications as these produce characteristic mass differences between the modified and unmodified peptides nuclear proteins do not resolve. Also, many spots may contain multiple, comigrating proteins (same mass, same isolectric point), which complicates identification and quantitation.
- Therefore, multi-dimensional liquid chromatography (LC) systems are being developed to replace 2D gels. Usually, the peptide mixture is first fractionated by ion-exchange chromatography and then the peptides in each fraction are further separated by reverse phase chromatography. Since peptide sequences can unambiguously identify proteins, tandem 2D LC-MS/MS procedures can be used with digests of unresolved protein mixtures, with the data analyzer unscrambling the information at the end to show what proteins were present in the original sample.
- Quantifying the differences in the levels of specific proteins between two samples can be very important. The relative staining intensity of corresponding spots on two 2D gels can be measured, or, if the proteins have been radiolabeled by exposure of the cells to a radioactive amino acid such as [35S]methionine before extraction, the radioactivity of the spots can be determined.
- However, MS can be used to give very accurate quantification simultaneously with identification using stable, heavy isotopes such as deuterium, which is one mass unit heavier than normal hydrogen, or 15N or 13C. If one sample is labeled by growing the cells with an amino acid containing normal 12C and the other with the amino acid containing 13C, the proteins from these samples will generate identical peptide mass fingerprints, except that one will be a precise number of mass units heavier than the other.
- The ratio of the two shows the difference in abundance. Proteins in extracts can also be post-labeled with normal and heavy **isotope-coded affinity tags (ICAT)**. One advantage of these methods is that the two samples can be mixed before analysis (cf DNA microarray analysis,as their mass spectra can

always be distinguished at the end. This avoids problems due to differences in sample loading and processing.

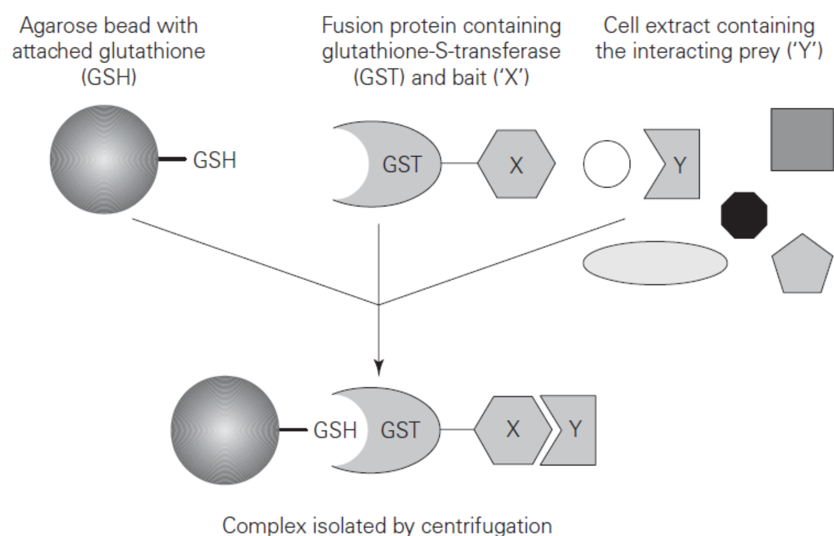


Alternative strategies for proteome analysis

➢ Protein–protein interactions

- Many proteins are involved in multiprotein complexes, and transient protein–Many proteins are involved in multiprotein complexes, and transient protein–protein interactions underlie many intracellular signaling systems.
- Thus, characterization of these interactions (the **interactome**) is crucial to our understanding of cell function. Such interactions may be stable, and survive extraction procedures, or they may be weak and only detectable inside cells. Stable interactions can be detected by **immunoprecipitation (IP)**.
- Here, an antibody raised against a particular protein antigen is added to a cell extract containing the antigen to form an **immune complex**. Next, insoluble agarose beads covalently linked to **protein A**, a protein with a high affinity for immunoglobulins, is added.
- The resulting **immunoprecipitate** is isolated by centrifugation and, after washing, the component proteins are analyzed by SDS gel electrophoresis. These will consist of the antigen, the antibody and any other protein in the extract that interacted stably with the antigen. These proteins can be identified by mass spectrometry.

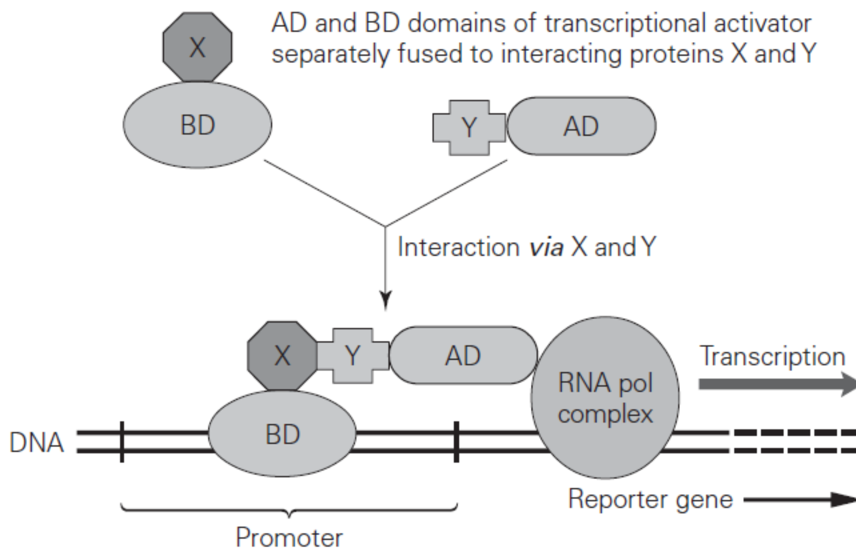
- The **pull-down assay** is a similar procedure. Here, the **'bait'** protein (the one for which interacting partners are sought) is added to the cell extract in the form of a recombinant fusion protein, where the fusion partner acts as an **affinity tag**. A commonly used fusion partner is the enzyme **glutathione-Stransferase (GST)**. Next, agarose beads containing immobilized **glutathione**, a tripeptide for which GST has a high affinity, are added.
- The GST-bait fusion protein binds to the beads, along with any other protein that interacts with the bait. The beads are then processed as for IP. This technique can be used for proteome-wide analysis. A collection of 6000 yeast clones is available with each clone expressing a different yeast protein as a GST-fusion expressed from a plasmid with a controllable promoter.
- The yeast **TAP-fusion** library is another collection where each yeast protein is tagged with an affinity tag called TAP and expressed, not from a plasmid, but from its normal chromosomal location following homologous recombination between the normal and the tagged version of the gene.
- The TAP tag is also an **epitope tag** (an amino acid sequence recognized by an antibody) so it not only allows purification of proteins that interact with each bait protein fused to the tag in an in vivo environment, but, since each tagged protein is expressed from its own chromosomal promoter, it also allows quantification of the natural abundance of every cellular protein by **immunofluorescence**.



GST pull-down assay for isolating proteins that interact with the bait protein 'X'.

Two-hybrid analysis:

- Two-hybrid analysis reveals protein–protein interactions that may not survive the rigors of protein extraction by detecting them inside living cells. Yeast has commonly been used to provide the in vivo environment.
- This procedure relies on the fact that many gene-specific transcription factors (transcriptional activators) are modular in nature and consist of two distinct domains – a DNA-binding domain (BD) that binds to a regulatory sequence upstream of a gene and an activation domain (AD) that activates transcription by interacting with the basal transcription complex and/or other proteins, including RNA polymerase II.
- Although normally covalently linked together, these domains will still activate transcription if they are brought into close proximity in some other way. Two hypothetically interacting proteins X and Y are expressed from plasmids as fusion proteins to each of the separate domains, that is BD–X and AD–Y.
- They are transfected and expressed in a yeast strain that carries the regulatory sequence recognized by the BD fused to a suitable reporter gene, for example β-galactosidase. If X and Y interact in vivo, then the AD and BD are brought together and activate transcription of the reporter gene. Cells or colonies expressing the reporter. gene are easily detected as they turn blue in the presence of X-gal, a chromogenic β-galactosidase substrate.
- The real power of this system lies in the detection of new interactions on a proteome-wide scale. Thus, if protein X is expressed as a BD–X fusion (the **bait**) and introduced into a yeast culture previously transfected with a cDNA library of AD–N fusions (the **prey**), where N represents the hundreds or thousands of proteins encoded by the cDNAs, any blue colonies arising after plating out the culture to separate the clones represent colonies expressing a protein that interacts with X.
- If the plasmids are isolated from these colonies and the specific AD–N cDNAs sequenced, the interacting proteins (N) can be identified. Although yeast is often used as the environment, the bait and prey can be from any organism. Bacteria and mammalian cells are also used to provide the in vivo setting for the interactions.



Principles of yeast two-hybrid analysis: activation of a reporter gene by reconstitution of active transcription factor from its two separated domains, activation domain (AD) and binding domain (BD).

- **Protein arrays:** By analogy with DNA microarrays, a **protein array** (or **protein chip**) consists of proteins, protein fragments, peptides or antibodies immobilized in a gridlike pattern on a miniaturized solid surface. The arrayed molecules are then used to screen for interactions in samples applied to the array. When antibodies are used, the array is effectively a large-scale enzyme-linked immunosorbent assay (ELISA) as used in clinical diagnostics and such arrays are likely to find future use in medicine to probe protein levels in samples. Protein arrays can also be screened with substrates to detect enzyme activities, or with DNA, drugs or other proteins to detect binding.