

gene function via the analysis of multi-pro

Affinity proteomic methods can translate genomic data into validated targets for drug discovery. Baits such as tagged gene product or small molecules obtained from cell-based assays are used to purify interacting proteins. These proteins are then identified by high sensitivity, high throughput mass spectrometric techniques. Successful examples of this novel method are discussed in this article, such as the discovery of key components of the prototypical NF- κ B inflammation pathway, which are now in drug screening programs.

**By Matthias Mann,
Protein Interaction
Laboratory (PIL),
University of
Southern Denmark**

Genome sequencing programs have now resulted in the blueprint for all possible protein drug targets. However, 'database mining' for interesting receptors, kinases and the like have often been only partially effective and sometimes frustrating exercises. It is becoming apparent that bioinformatic techniques alone do not offer sufficient validation and justification for starting a drug-screening program. A number of 'functional genomics' approaches are now available to study the human genes in a large-scale format. Among these, proteomic methods¹ promise to occupy a key role because proteomics allows the direct, large-scale study of proteins, which are the targets for the majority of drugs. Proteomics has previously been largely synonymous with differential 2D gel approaches, that is, measurement of protein expression changes as they appear on high-resolution 2D gels².

More recently, it has been realised that proteomics can be used in clever, scaled-up biochemical strategies, which can lead much more directly to biological function. Here I will describe an approach developed in our laboratories to place a protein into a functional context through its interacting proteins. This 'affinity proteomic' approach allows the validation of drug targets derived from searches of the human genome. The genes found differentially regu-

lated in cDNA array experiments can also be validated in this way.

Defining multi-protein complexes by affinity purification and mass spectrometry

The central idea of our strategy is to isolate a multi-protein complex by biochemical means, separate it into its components, identify the constituents by mass spectrometry and database searching and finally to verify the bona fide role of the components found in the complex³. Such a strategy can elucidate the function of novel genes through their interaction partners and it can also be used in a comprehensive way to obtain a protein interaction map of the cell. We can combine affinity purification approaches with high-throughput mass spectrometric identification of proteins for the elucidation of the function of disease genes, the discovery of drug targets and the mapping protein interactions in pathogens.

Purification of the protein complex

To study a particular complex by proteomics, we first need a 'hook' or affinity tag for biochemical purification (Figure 1). There are three generic ways to obtain such 'hooks' for affinity purification. Antibodies can

the mass spectrometric protein complexes

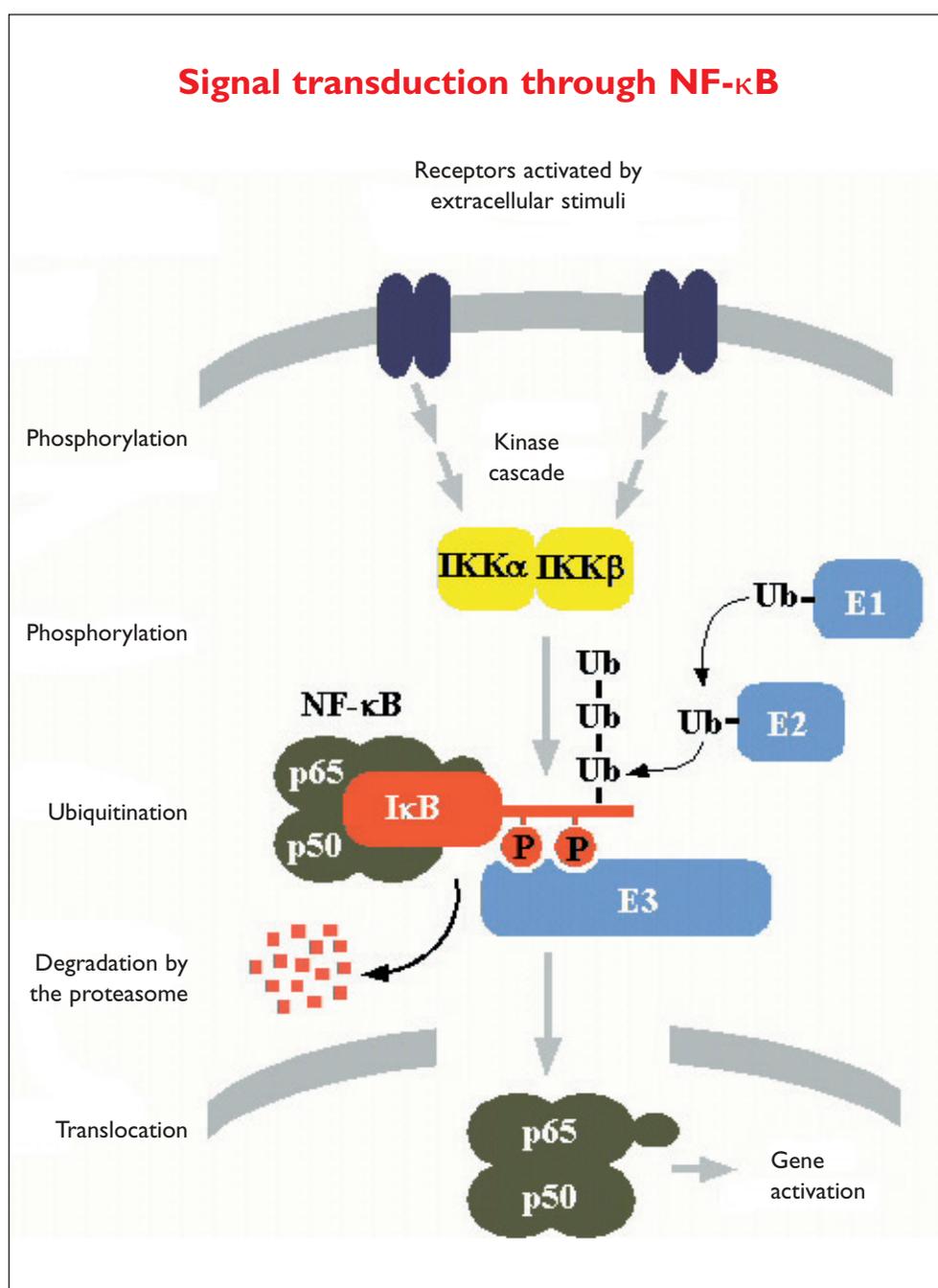


Figure 2

The NF- κ B pathway. Affinity proteomic methods determined two key points in the cascade. The IKKs are the kinases that phosphorylated I κ B which in turn shields the nuclear localisation sequence of NF- κ B. After phosphorylation, I κ B is recognised by an E3 ligase, which, together with the E2 and E1 subunits of the ubiquitination machinery, attach multiple copies of ubiquitin to I κ B. This leads to destruction of I κ B and gene activation by NF- κ B in the nucleus. Both the IKKs and the E3 receptor are potential drug targets and have been identified by the affinity proteomic methods described in this article

Analysis

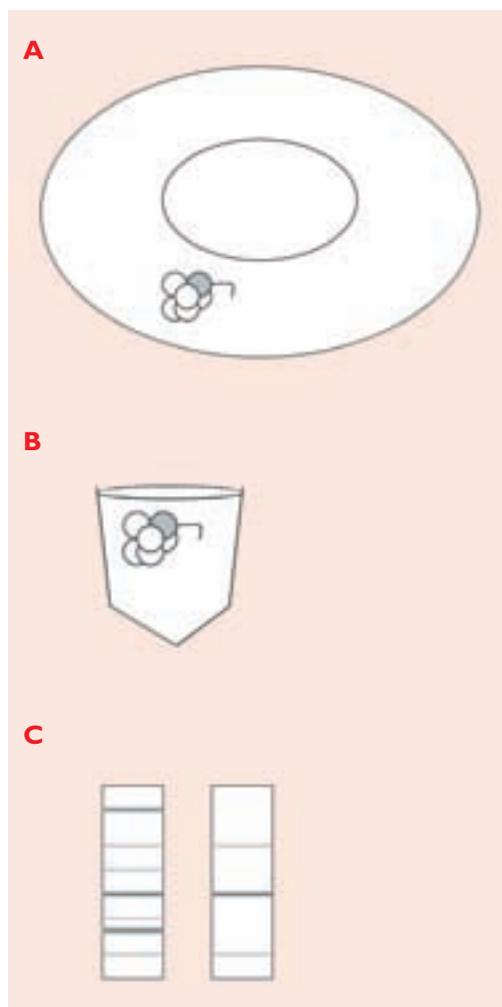
Figure 1

Schematics of protein complex characterisation by mass spectrometric methods.

A A multi-protein complex is shown in the cytosol of a cell. One member of the complex (gray) has been fused to a short epitope – the 'hook'.

B The 'hook' is used for purifying the protein complex from cell extract by antibodies which recognise the epitope.

C The bound protein complex is specifically eluted using the properties of the 'hook' and separated by electrophoresis. Bands which are different between the bait (left lane) and a control are excised and identified by mass spectrometry



be used to precipitate the complex from a cell extract. Antibody precipitation captures endogenous levels of the protein complex and is, therefore, the method most likely to deliver its *in vivo* state. Phage display antibody libraries, other large collections of antibodies or streamlined methods of obtaining antibodies are advantageous when purifying complexes in this way. An alternative method is to construct a fusion of the cDNA of a member of the complex and an epitope tag against which a defined antibody exists. This construct is then expressed in target cells.

The affinity tag provides a generic 'handle' with which to retrieve the protein together with associated binding partners. A third variation on the theme is to express an affinity tagged version of a member of the complex in a convenient system such as *E. coli*. Subsequently the expressed protein is immobilised on beads followed by incubating the beads with cell extract. The protein and its binding partners are then

eluted from the beads. The latter strategy is currently the most 'scalable'. However, great care has to be taken concerning the design of controls and follow up experiments to verify the interactions. In addition to the affinity purification schemes mentioned above, non-protein baits can be used, such as small molecules that have specific but unknown protein binding partners. In this way, molecules that registered as 'hits' in cellular assays can be connected to cellular target proteins.

The strategies outlined above can be used as one-step procedures or may be combined with several conventional biochemical separation methods, such as gradient centrifugation and various forms of chromatography. In this way protein complexes can be obtained specifically for different cell compartments.

In the immunoprecipitation or 'pull down' step, non-specifically binding proteins as well as physiological interactors are bound to the target and the beads. A balance must be struck between minimising unspecific protein background by stringent washing and, at the same time, preserving weak interactions. A promising advance in this area has been the development of alternative tagging systems that allow specific retrieval with little background. These involve elution by proteolytic cleavage of a specific sequence inserted between the tag and the bait, double tags or a combination of these⁴.

Separation of bound proteins and preparation for analysis

Following elution, the components of a protein complex are separated by gel electrophoresis. Usually, the complexity of the mixture is relatively low, often allowing 1D rather than 2D gel analysis. In our experience, protein mixtures of up to 100 members can often be analysed by 1D gels since mass spectrometric methods can now easily resolve the identities of co-migrating proteins (see below). One-dimensional SDS gels have added advantages over 2D gels as almost all proteins can be visualised, 1D gels are easy to use and 1D gel experiments can easily be scaled up to large numbers.

After staining – usually by silver – proteins are excised from gels either manually or with the help of a robotic spot picker. Subsequent analysis is also either manual or in an automated sample handling and workflow system. The latter has the advantage of avoiding errors due to manual tracking of spot, batch and spectrum identity. Proteins need to be reduced to peptides in order to be analysed by mass spectrometry. In our laboratory, protein spots are enzymatically degraded in a streamlined system using 96 well plates in order to accommodate the large numbers of pro-

Analysis

teins generated by the many pull down experiments involved in large-scale protein interaction mapping.

Mass spectrometric analysis of peptide mixtures

The resulting peptide mixtures must then be analysed by mass spectrometry. This technique has been revolutionised in recent years due to the introduction of new ionisation methods (electrospray and MALDI, or matrix assisted laser desorption/ionisation, mass spectrometry), streamlined handling techniques and novel mass spectrometric instrumentation and data analysis. It is these advances that much of the promise of proteomics is built on.

As a first screen, peptide mixtures are subjected to MALDI mass fingerprinting. Large numbers of aliquots containing peptide mixtures are spotted on targets which are then inserted into mass spectrometers. They are then automatically screened by directing laser shots on each spotted position, resulting in a mass spectrum of the peptides at that position. Software then screens this list of peptide masses or 'mass fingerprint' against large sequence databases, calculating the theoretically predicted list of peptide masses and revealing the identity of a large percentage of the proteins.

Protein bands that are very faint (low silver staining or low nanogram level), contain complex protein mixtures or those proteins that are represented only partially in a database can often not be identified conclusively by MALDI mass fingerprinting and need to be sequenced in a second experiment. Mass spectrometric sequencing of peptides is performed with electrospray ionisation. Peptides are passed through very fine needles or capillary columns the end of which is at high electrical potential. This leads to electrical dispersal (electrospray) of the liquid, liberating the peptides into the gas phase from which they are drawn into the mass spectrometer.

The peptides are isolated in the first part of the mass spectrometer, fragmented by collision with background gas atoms and the fragments mass measured by the second part of the mass spectrometer (tandem mass spectrometry). Many peptides can be fragmented in the same experiment by selecting different peptides in the first part of the mass spectrometer. Current instrumentation (so-called 'quadruple time-of-flight instruments') allow characterisation of minute quantities of peptides at relatively high throughput. The results of the fragmentation are partial amino acid sequences gained from the mass differences of fragment ions.

Sophisticated database searching software has been developed that will screen this data against amino acid and nucleotide databases. Recently it has become pos-

sible to search directly against the draft sequence of the human genome allowing the identification of virtually every human protein.

Examples of complexes analysed by mass spectrometry

We have analysed a large number of protein complexes by the methods sketched above. Examples include FLICE/Caspase-8, a key member of the apoptotic signaling pathway⁵ and telomerase⁶, a key molecule in cancer and ageing, which is now subject to intense evaluation as an adjunct to cancer therapies. The yeast and human spliceosome⁷ have been analysed as well as the nuclear pore complex⁸. In each case, novel members were found, placing them into a functional context and shedding light on the function of the complex. Complete compartments of the cell can also be mapped, leading to a categorisation of the both the compartments and the proteins which have been found.

Comparison to two-hybrid system

The yeast two-hybrid system is similar to the proteomic study of multi-protein complexes as it also attempts to obtain biological function through protein-protein interaction. This screen is based on the binary interaction of bait and prey species in the cell nucleus and furthermore does not reveal interactions that depend on post-translational modifications. For these reasons, there is relatively little overlap between protein-protein interactions determined by the two-hybrid system and the affinity proteomic approach described here. For example, five new members of the U1 small nuclear ribonucleoparticle, a component of the spliceosome, were revealed by mass spectrometry but not by large scale two-hybrid approaches. On the other hand, the two-hybrid approach is readily automated and can be a good complement to the methods described here.

The NF- κ B pathway

Investigation of the NF- κ B pathway, which we undertook in collaboration with Signal Pharmaceuticals and the Medical School of Jerusalem, illustrates the power of affinity proteomic approaches. The NF- κ B pathway is important in inflammation and is also a prototypical signaling pathway (Figure 2).

Briefly, signals transmitted through cell surface receptors activates a kinase cascade which eventually leads to translocation of the transcription factor NF- κ B to the nucleus where it turns on a multitude of genes involved in inflammation response. Two key players of this pathway have been elucidated by affinity proteomics: The kinases that phosphorylate the I κ Bs proteins that shield the nuclear localisation

Analysis

References

- 1 Pandey, 2000 #925.
- 2 Hochstrasser, 1997 #691.
- 3 Lamond, 1997 #686.
- 4 Rigaut, 1999 #783
- 5 Muzio, 1996 #607.
- 6 Lingner, 1997 #709.
- 7 Neubauer, 1997 #656;
Neubauer, 1998 #681.
- 8 Rout, 2000 #905.
- 9 Mercurio, 1997 #739.
- 10 Yaron, 1998 #768.

sequence of NF- κ B were purified and sequenced by mass spectrometry leading to their cloning and characterisation⁹.

These proteins, which are now called IKKs, are in drug screening programs at several companies. A second major question concerned the identity of the receptor recognising the phosphorylated sequence stretch of the I κ Bs thereby conveys specificity to this signalling pathway. Binding of the receptor leads to ubiquitination of I κ B which in turn leads to its destruction by the proteasome machinery and the 'liberation' of NF- κ B which can then translocate to the nucleus.

Using antibodies against NF- κ B, the complex was assembled in stimulated and unstimulated cells. The assembled complexes were then incubated with the phosphopeptides corresponding to the phosphorylated sequence on I κ B. Profiles of the eluted proteins showed subtle differences between experiments and control. The bands corresponding to these differences were excised, sequenced by mass spectrometry and the results searched in sequence databases. This led to the identification of a drosophila and human homolog of this novel protein which were used to clone the protein¹⁰. The interaction found here has turned out to be of key significance not only in this signalling pathway but also in the interaction modified in colon cancer.

Apart from demonstrating the power of affinity proteomic techniques, the above example also highlights the need for very sophisticated analytical capabilities when performing these experiments.

- Binding proteins were retrieved in very small quantities and some bands had to be identified from amounts that were not even silver stainable. This necessitated analytical sensitivity more than hundred

fold higher than available until recently.

● Proteins were found in mixtures and some of the bands contained more than five gene products. Without sophisticated database searching capability, these bands, or at least the novel components would not have been identified.

● A combination of directed strategy and generic, large-scale analysis capability was needed to solve this problem. The interactions in question would not have been picked up by a completely generic protein-protein interaction screen nor would they have been accessible to a less streamlined analytical system.

Conclusion

Advances in the mass spectrometric characterisation of proteins and the availability of all human genes in sequence databases now make it possible to assign the function of genes at the protein level. These methods are much more direct than genetic methods in that they take post-translational modification, localisation in the cell and other attributes of mature proteins into account. A number of interesting complexes and potential drug targets have already been identified. As proteomics technology advances affinity proteomics methods will become increasingly powerful and may become methods of choice to define the function of genes 'mined' from the genome. DDW

Matthias Mann works at the Protein Interaction Laboratory (PIL), University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark; mann@cebi.sdu.dk and Protana A/S, Staermosegaardsvej 16, DK-5230 Odense M, www.protana.com

ADVERTISERS' INDEX

Amersham Pharmacia plc	39	Inpharmatica Ltd	inside back cover
Arqule Inc	outside back cover	MediChem Research Inc	12
Bruker Daltonik GmbH	63	Micromass Ltd	42
Discovery Technologies Ltd	17	Molecular Devices Corporation	37
DoubleTwist Inc	31	PE Biosystems (Perceptive Biosystems)	6
Gene Logic Inc	8	PE Biosystems (PE Sciex)	65
Genetics Computer Group	24	SignalGene Inc	23
Gilson Inc	inside front cover	Triplos Inc	40
Incyte Genomics Inc	5	Zinsser Analytic	44
Informax Inc	26		