Topological proteomics: a new approach to drug discovery

Joanna Owens, joanna.owens@drugdiscoverytoday.com

A new sensitive proteomics technology has been developed that can identify, localize and characterize entire protein networks within a single cell, which could prove invaluable in identifying new drug targets and selecting potential lead compounds. MelTec (Magdeburg, Germany) uses robotic imaging technology that analyzes thousands of cells simultaneously for an almost unlimited number of proteins. The proprietary technology, Multi-Epitope-Ligand Kartographie (MELK), enables the identification of proteins and protein clusters at the subcellular level by integrating cell biology and biomathematical algorithms developed by MelTec.

Protein topology

Walter Schubert, CEO of MelTec, explains the need for studying functional protein networks of the entire cell: 'Obviously, the cell does not care about all of the thousands of proteins that it could express at a time. Rather, it codistributes particular proteins within a cell into networks that will exert a specific function, such as migrating from A to B.' Therefore, every cell is topologically determined in a precise manner so that proteins that exert particular functions can interact with each other. For this reason, the most powerful proteomics technology will be one that takes a 'snapshot' of the protein fingerprint of a cell at the moment when the proteins interact to exert a function (Fig. 1). During MELK's initial stages of development, it was discovered that mature vascular cells in humans transdifferentiate to muscle stem-cells *in vivo*[1], (a mechanism later confirmed by experimental procedures[2]), providing proof-of-concept early on.

Schubert says that the gain of information using the MELK technology is enormous because it analyzes information from 2000 subcellular compartments. Therefore, analysis of 1×10^6 cells gives an information gain of 2×10^9 compared with conventional proteomics technologies. In addition to identifying key proteins in biological pathways, the technology can also be used to determine hierarchies of proteins within networks. 'This might lead us to identify those proteins that switch on pathways, which are of primary importance because these are excellent drug targets,' comments Schubert, 'So, what we do is a kind of target selection procedure that can be directly linked with drug selection at the single-cell level, to come up with much more specific target proteins.'

How are the proteins identified?

Proteins are identified and tracked in the cell using large protein libraries. Meltec has developed specific protocols based on a highly stable biophysical diffusion principle, so that applied proteins from a library, such as antibodies, can interact with a protein anywhere within, or on the surface of, a cell. This biophysical interaction, which can be detected as a phototonic signal, is highly stable, so that every signal detected indicates an interaction between a library protein and an unknown. It is, therefore, highly reproducible.

This cell biology approach is integrated with new pattern-recognition algorithms developed by the company. These are biology-driven algorithms, that is, when a cell produces a specific protein pattern to exert a specific function, the algorithm learns from the biological pattern found. Schubert explains: 'We detect it, we primarily do not understand the pattern, but nevertheless we put it in the memory store, and by matching different biological situations the algorithm learns to recognize specific patterns that encode a cellular function. The whole system is a

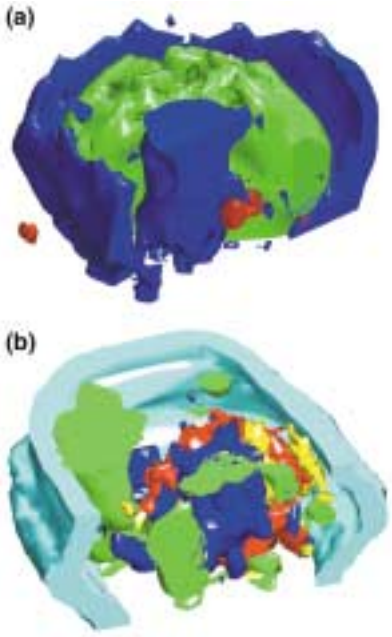

Figure 1. Topological proteomic fingerprints. Using MelTec's Multi-Epitope-Ligand Kartographie (MELK) technology to visualize the three-dimensional distribution of 40 different protein species in lymphocytes to show significant differences in the protein distribution between a normal and diseased immune cell. **(a)** A normal lymphocyte; the proteins reach the cell's surface. **(b)** A diseased lymphoma cell; the proteins do not reach the cell's surface, they are stored in the endoplasmic reticulum. Reference proteins indicate the presence of Golgi-cisternae (red) and *trans*-Golgi vesicles (yellow).

1359-6446/01/$ – see front matter PII: S1359-6446(01)01995-X

database that more or less has a couple of different levels of functional recognition systems.' He continues, 'The most important thing is that you come up with a colocalization pattern; the next thing is to interpret this by matching different cell types for the same proteins and learning the selectivity of biological patterns that are recognized.'

The MELK technology can be integrated with existing proteomics tools. For example, a two-dimensional electrophoresis gel could detect proteins within a disease pathway that are specifically regulated or modified under certain conditions. The MELK technology could then be used to look at the functionality of such proteins within a network that has already been characterized. This enables further validation of proteins identified using traditional proteomics technology as potential drug targets.

Schubert also explains that MELK technology can enhance our use of genomics data because the functional analysis of proteins encoded by the genome will reveal how genes are organized into functional clusters. Studying the regulation of large gene clusters could lead to new targets for gene therapy and drug discovery. 'Using MELK will allow this to be done at the single cell level,' Schubert comments, 'For example, a T cell has different patterns of gene expression to a muscle cell, and although these genes are present in all cell types, we can say which genes belong together and which definitely do not. This enables us to understand how a genome specifically works in a single cell type.'

As with traditional proteomics technology, the main limitation is the number of machines needed to analyze the whole proteome of all cell types. The technology is currently capable of analyzing 40,000 cells per day per machine. However, Schubert points out that the patterns of expression are so disease-specific that the technique is highly robust. For example, if you study prostate cancer and you find an identical pattern in five cases that are not present for fifty other diseases, then you would not need to confirm this by using a larger data set. Schubert comments, 'This is a question of scaling up the technology to decipher 5–8 different cell types that are relevant for clinical diseases in the next five years is something that is possible. However, we do not know yet how long it will take to cover the whole functional protein network information of all cell types in humans. This is something that is highly complex, and needs definite integration of all available technologies.

Targets for ALS and cancer

Amyotrophic lateral sclerosis (ALS; also known as Lou Gehrig's disease) is a fatal, progressive neurodegenerative disease that causes two deaths in every 100,000 people in the USA, and accounts for a higher mortality than Huntington's disease and multiple sclerosis. MelTec has used their proteomics technology to identify the first biomarker for ALS that is detected in mononuclear blood cells. This enables screening for ALS to be done using small samples of blood. The biomarker is a cell surface protein and provided the basis for protein fingerprint studies that revealed a specific linkage of proteins in a particularly unusual form that is not seen in healthy individuals or other diseases. The team has developed a way of blocking this linkage using a recombinant protein, and will shortly begin preclinical evaluation of this candidate drug.

Similarly, the group has identified another protein cluster that is necessary for tumour cells to polarize, which enables them to migrate. This is an essential function for all tumours to be able to metastasize, and this is the first time the mechanism by which a tumour cell achieves migration has been determined. The protein cluster identified was found to be highly specific for the tumour type, and MelTec are now studying other tumour cells to find similar protein patterns.

Other applications

MelTec is currently collaborating with Evotec OAI (Hamburg, Germany) to apply this technology to ADME/Tox studies. This is based on using the specific protein fingerprint to identify sensitivity and toxicity to drugs. The advantage is that the effect of a specific drug can be studied for a specific functional pathway within a cell before cell death occurs. Schubert explains, 'For example, a liver cell has a specific proteomic fingerprint, and MelTec is establishing an assay to test, in a high throughput manner, any drug that could influence liver-cell-specific protein fingerprints. To be able to select out drugs with undesirable toxicology profiles at a very early timepoint will obviously be very helpful.'

In addition to the biomarker identified for ALS, Schubert believes that this technology can be used to develop biomarkers for other diseases: 'It could be that every disease shows a unique cellular fingerprint in the immune system, or in any other cell type, that we can easily detect. We do not know yet about other diseases, but we are working on that.' The detailed information about cell surface proteins could also be used for drug and vaccine targeting. For example, endothelial cells express different protein patterns depending on the organ or tissue to which they belong. This provides an opportunity for the specific targeting of drugs and vaccines to organs or tissues.

The topological database that MelTec is developing will take some time to complete, but the company intends to make it available for researchers to benefit from the wealth of information it will eventually contain.

References

1 Schubert, W. (1992) Antigenic determinants of T lymphocyte α β receptor and other leukocyte surface proteins as differential markers of skeletal muscle regeneration: detection of spatially and timely restricted patterns by MAM microscopy. *Eur. J. Cell Biol.* 58, 395–410

2 De Angelis, L. *et al.* (1999) Skeletal myogenic progenitors originating from embryonic dorsal aorta coexpress endothelial and myogenic markers and contribute to postnatal muscle growth and regeneration. *J. Cell Biol.* 147, 869–878