The study of proteome

ASSIGN BUSTER

The breakthrough in the study of genomics has created a great impact for biological research. For instance, the complete genome sequencing accomplished for various model organisms such as bacteria, yeast and drosophila has been a key to understanding the diversity and functions in living system (Liebler, 2002). However, the gene products, proteins are playing crucial complex functions in sustaining the living system. Therefore, in order to acquire a clearer picture about the function of the living cells at molecular level, it is important to study the entire set of protein expressed by a genome, cell, tissue or organism throughout the cell cycle as known as proteome (Monti *et al.* , 2005).

This study of proteome is termed proteomics which is mainly focus on the identification of proteins, determination of the interaction between various proteins as a part of a larger network and outlining the 3D structures of proteins (Liu & Hsu, 2004). According to Liebler (2002), the applications of proteomics comprised of mining, protein-expression profiling, protein-network mapping and mapping of protein modifications. These four principals then bring about the branches of proteomics which are the sequence and structural proteomics, expression proteomics, interaction proteomics as well as functional proteomics.

i. Sequence and structural proteomics

Structural proteomics is generally aimed at efficiently convert the gene sequence into 3D structural information which will then being used to facilitate the prediction of protein folding and functions afterward (Yee *et al.* , 2002). 3D structure prediction and determination of proteins on a genome-wide scale are managed to provide substantial information in understanding

the structure-function relationships of the particular proteins (Liu & Hsu, 2004).

The approaches commonly employed for this area of study are NMR (nuclear magnetic resonance) spectroscopy and X-ray crystallography. In a study done by Yee *et al.* (2002), structural proteomics of small proteins based on data of 513 proteins from five microorganisms is done by using NMR spectroscopy. Small proteins are more acquiescent for structure analysis primarily due to its solubility as protein aggregation appeared to be a limitation for structural proteomics (Yee *et al.* , 2002). NMR evaluation of proteins is performed after ran through a series of cloning, expression and purification processes.

The chosen targets for NMR screening by Yee *et al.* (2002) were single chain polypeptides with not more than 23 kDa molecular mass. PCR was involved to amplify the targets from genomic DNA. Subsequent cloning of the target proteins into vectors was then carried out. Batch Ni^{2+} affinity chromatography was done for the extraction of cells (Yee *et al.* , 2002). Varian INOVA 500- or 600-MHz spectrometer was used to attain all the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra and the NMRPIPE software package was utilized to processed the data (Yee *et al.* , 2002). The HSQC spectrum is to provide a diagnostic fingerprint of a protein.

Twelve of the proteins were considered to be suitable for structure determination and have their 3D structures successfully analysed which were conserved without functional annotation (Yee *et al.* , 2002). Among them, eight out of twelve were the members of uncharacterized cluster of

orthologous groups (COG) or protein superfamilies. In order to know whether a 3D structure can be predicted, the sequences were submitted to SwissModeler to be analysed based on sequence similarity.

ii. Expression proteomics

Human Protein Atlas (HPA) launched by Swedish researchers in 2003 is a comprehensive effort in mapping the proteins expressed in the body's tissue and cells according to the genes documented by Human Genome Project (Marx, 2014). HPA principally looked at the spatial distribution and protein-expression patterns of human proteins as well as comparing the protein expression between cancerous and normal tissues. The expression proteomics is a focus for proteome study which targets on measurement of the up- and down-regulation of protein levels as well to explore the protein expression patterns in abnormal cells for biomedical application (Monti *et al.* , 2005).

Approaches like mass spectrometry, 2D-gel electrophoresis and protein arrays are used in expression proteomics. Mazzanti *et al.* (2006) as done an investigation upon the differential expression proteomics of human colon cancer to look at the operative pathway of early stages of human colon cancer and to compare with normal colon tissue. In this case, they were using 2D-gel electrophoresis to observe the differences in protein expression at various differentiation levels.

After the tissues were properly collected and homogenized, Western Blot analysis was carried out to evaluate the expression levels of the proteins and followed by 2D-GE. The differences in intensity of each spot and the

normalized spot volume were calculated and statistically analysed. Matrix-assisted laser desorption ionization-time-of-flight mass spectroscopy analysis of tryptic peptides was performed to detect the differentially expressed samples processed by 2D-GE (Mazzanti *et al.* , 2006). Lastly, a search through NCBI non-redundant databases and SWISSPROT was completed to identify the proteins.

As a result, several proteins were found differentially expressed in colon cancer and normal tissue. Mazzanti *et al.* (2006) have obtained 11 and 15 proteins from healthy and cancerous tissues which were categorized according to the pathway they involved respectively. In addition, they also claimed that cancer tissues preferably utilize glycolysis due to selective repression of β-subunit ATPase expression and limited substrate availability to mitochondria (Mazzanti *et al.* , 2006).

iii. Interaction proteomics

Interaction proteomics emphasis on the elucidation of protein-protein interaction to which the multitude functions of proteins in association with individual components of proteome and the formation of multimeric complexes (Völkel *et al.* , 2010). Frequently utilized approaches for interaction proteomics are like mass spectrometry (MS), affinity captured combined with mass spectrometry and yeast 2-hybrid techniques.

Völkel *et al.* (2010) mentioned that characterize protein complexes from bacteria, yeast, mammalian cells and some multicellular organisms can be done by using tandem affinity purification-mass spectrometry (TAP-MS). TAP is functioned based on the dual purification of protein assemblies by

sequential utilization of two affinity tags spaced by a TEV (tobacco etch virus) protease cleavage site (Völkel *et al.,* 2010). TEV is a sequence-specific protease that cleaves a recognition site of seven-amino-acid (Glu-X-X-Tyr-X-Gln/Ser) which is rarely found in mammalian proteins (Berggård *et al.,* 2007; Völkel *et al.,* 2010). This epitope-tagging purification strategy further improves the sample purification prior to protein identification by MS.

In the first purification step, the TAP-tag protein is bound to specific column like immobilized Ig (Berggård *et al.,* 2007; Völkel *et al.,* 2010). Retrieval of TAP-tag protein is done by adding TEV-proteinase. The protein complex is then immobilized by CBP (calmodium-binding peptide) of TAP tag to a second column, the calmodulin-coated beads in the second affinity step. This calcium-dependent CBP-calmodium interaction enables the elution of final protein complex for MS protein identification by means of calcium chelation (Völkel *et al.,* 2010).

The TAP purification of protein complex is independent of the availability of specific antibodies. It is used to accomplish large-scale protein interaction mapping of lower organisms and decode the smaller interactomes and signalling pathway in mammals (Völkel *et al.,* 2010). However, this technique may not be able to identify low stoichiometric complexes, transient interactions or under-represented interactions occurred specifically in some physiological states of cells growing exponentially (Völkel *et al.,* 2010).

iv. Functional proteomics

The objectives of functional proteomics are to characterize the biological function of unknown proteins, multiprotein complexes and descript the

cellular mechanisms as well as signalling pathway at molecular level (Monti *et al.* , 2005). Henning and Ilag (2003) stated that the protein functions assessment could be done based on the individual cell types and interaction of protein partners. Functional proteomics implement the concept that functional inactivation within the physiological environment of a targeted protein can be used to determine its function (Henning & Ilag, 2003).

Tools that are typically employed in functional proteomics include 2D-gel electrophoresis, MS, MALDI-TOF (matrix-assisted laser desorption ionization-time of flight) MS and affinity-based procedure. The affinity-based procedure is worked on the idea that a suitable tag of the expressed target protein can be used as a bait to fish out its partners from a cellular extract (Monti *et al.* , 2005). Brookes *et al.* (2002) utilized a high-throughput two-dimensional blue-native electrophoresis in the study of functional proteomics for mitochondria and signalling pathway. As mitochondrial membrane proteins exhibit hydrophobic properties, precipitation is occurred in standard 2D gels during the first dimension isoelectric focusing. Thus, the 2D blue-native gel electrophoresis is used resolve this problem.

The first dimension of 2D blue native gel electrophoresis involved the solubilization the membrane protein complexes into its native form. This is followed by the second dimension utilizing the sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-page) with its gel staining performed by employing a mixture of Coomassie blue G-250 and R-250. SDS-page is worked on denaturing the protein complexes and enables them to resolve into their component subunits (Brookes *et al.* , 2002). A partial

mitochondrial proteome map was then assembled by using MALDI-TOF MS coupled with tryptic peptide fingerprinting.

This 2D blue native gel electrophoresis is found suitable for proteomic analysis of mitochondria proteins in which it could separate large amount of mitochondria proteins and large proportion of respiratory chain complexes as presented as the spots on the gel can be readily identified according to their position (Brookes *et al.* , 2002). The functional associations between different respiratory complexes in mitochondria like cytochrome *C* , respiratory complexes III and IV can be interpreted from the gels. From this study, Brookes *et al.* (2002) suggested that the respiratory complexes are able to assemble into sub-complexes with limited functionality regardless of the presence of mtDNA encoded subunits.