CHAPTER I

INTRODUCTION

1.1 Principle and rationale

Cancer is a major public health problem in the United States and many other parts of the world. One in 4 deaths in the United States is due to cancer. In term of mortality rate, lung cancer is the major cause of death among cancers with incidence of 87,260 cases in males and 72,220 cases in females in the United States in 2013 [1]. Smoking is the predominant risk factor for lung cancer. It is directly linked to lung cancer in 90% of women and 79% of men [2]. Although a substantial number of lung cancer patients have the potential to be successfully cured by early treatment, early diagnosis and prognosis of patients with lung cancer are still poor and responsible for the low overall 5-year survival rate of only 14% [3]. A key challenge in cancer medicine is to detect the disease as early as possible for lung cancer patients to receive appropriate and potentially curative treatments.

Recent studies have suggested that the abnormal glycosylation may be the indicator of cancer development [4]. Protein glycosylation is one of the most abundant structural forms of post-translational modifications (PTMs) of proteins and it plays a significant role in cellular differentiation, adhesion, immunological recognition or response, proliferation, invasion, metastasis and angiogenesis [5-8]. The production of various aberrant glycosylated proteins which occur through genetic variations at glycosylation sites or alternation of the glycan structures of proteins have been shown to correlate with the development or cancer progression compared to the normal cells [9]. Therefore, the abnormal glycoproteins are useful as a potential biomarker and have been utilized for prediction or early diagnosis of cancer [10].

1.2 Lung cancer

Lung cancer is the leading cause of cancer–related deaths in the worldwide. Recently, the American Cancer Society estimated the total numbers of new cases and deaths to be approximately 28% and 26% in the United State in 2013 [1]. However, global statistics estimated that 28% of all male cancer deaths and 26% of all female cancer deaths were found in the United States in 2013 **(Figure 1.1)**. In Thailand, lung cancer is the second most common cancer in males after liver cancer and the fourth in females after cervix, breast and liver cancers. There is a higher incidence rate of lung cancer in northern Thailand than other areas [11]. In 2005, there were 535 new cases of lung cancer diagnosed in 326 cases of males and cases of 209 females in Northern Thailand.

Estimated New Cases*

Males			Females		
Prostate	238,590	28%	Breast	232,340	29%
Lung & bronchus	118,080	14%	Lung & bronchus	110,110	14%
Colorectum	73,680	9%	Colorectum	69,140	9%
Urinary bladder	54,610	6%	Uterine corpus	49,560	6%
Melanoma of the skin	45,060	5%	Thyroid	45,310	6%
Kidney & renal pelvis	40,430	5%	Non-Hodgkin lymphoma	32,140	4%
Non-Hodgkin lymphoma	37,600	4%	Melanoma of the skin	31,630	4%
Oral cavity & pharynx	29,620	3%	Kidney & renal pelvis	24,720	3%
Leukemia	27,880	3%	Pancreas	22,480	3%
Pancreas	22,740	3%	Ovary	22,240	3%
All Sites	**854,790**	**100%**	**All Sites**	**805,500**	**100%**

Estimated Deaths

Males			Females		
Lung & bronchus	87,260	28%	Lung & bronchus	72,220	26%
Prostate	29,720	10%	Breast	39,620	14%
Colorectum	26,300	9%	Colorectum	24,530	9%
Pancreas	19,480	6%	Pancreas	18,980	7%
Liver & intrahepatic bile duct	14,890	5%	Ovary	14,030	5%
Leukemia	13,660	4%	Leukemia	10,060	4%
Esophagus	12,220	4%	Non-Hodgkin lymphoma	8,430	3%
Urinary bladder	10,820	4%	Uterine corpus	8,190	3%
Non-Hodgkin lymphoma	10,590	3%	Liver & intrahepatic bile duct	6,780	2%
Kidney & renal pelvis	8,780	3%	Brain & other nervous system	6,150	2%
All Sites	**306,920**	**100%**	**All Sites**	**273,430**	**100%**

Figure 1.1 Statistics of the ten types of leading cancer in new cases and deaths in United States estimated in 2013

Cigarette smoking is the main risk of lung cancer with 10 to 20 fold increased risk of lung cancer in smokers compared with non-smokers [12]. To date, epidemiological studies have identified several environmental, genetic, hormonal and viral factors associated with lung cancer risk including outdoor air pollutants, previous lung diseases and dietary factors [13-15].

Lung cancer can typically be grouped into two large categories: small-cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), which account for 15% and 85% of lung cancers, respectively. NSCLC consists of three major histological subtypes, adenocarcinoma, squamous cell carcinoma and large cell carcinoma [16]. Last few decades, SCLC and squamous cell carcinoma subtypes of NSCLC were frequently found histological subtypes of lung cancer in the initial period of the smoking more than adenocarcinoma subtypes of NSCLC. Recent studies have reported that adenocarcinoma is the most common histological type of lung cancer whereas the incidence of SCLC is decreased. It may be explained by the lower percentage of smokers and by the change in cigarette composition [17]. While adenocarcinoma was the most common form of lung cancer in young patients of which there was disproportionately high ratio of non-smoker [18, 19]. However, because of the poor early detection of lung cancer, even under the most advanced treatment strategies, about 85% of lung cancer patients die within 5 years of diagnosis [20, 21]. Therefore, early detection of lung cancer is crucial and beneficial for patients to receive appropriate and curative treatment, which may result in a longer survival period.

1.2.1 Types of lung cancer

It is well known that lung cancers can be categorized into two major types based on the characteristics of the disease and its response to treatment [22].

1.2.1.1 Non small cell lung cancer (NSCLC)

NSCLC are the most common lung cancer, accounting for about 80% of all cells found in the tumor. There are three subtypes of NSCLC as follows:

Adenocarcinoma: Adenocarcinoma (AC) usually arises from the mucus producing cells in the lung. Most cases of adenocarcinoma are associated with smoking

more than other types of cancers, but this is also observed as well in lung cancer of non-smoking patients. About two-thirds of adenocarcinomas develop in the outer regions of the lung, while one-third develops in the center of lung. The course of this cancer varies widely. Because it develops slowly and causes few or no symptoms until it is far advanced. In some cases, however, it can be extremely aggressive and rapidly fatal. In 50% of cases in which this cancer spreads, it spreads only to the brain. Other common locations it spreads to include the other lung, the liver, the adrenal glands, and bone. It accounts for approximately 40% of all lung cancer and it is more commonly found in women than in men.

Squamous cell carcinoma: Squamous cell carcinoma (SCC) is usually found in the center of the lung, either in a major lobe or in one of the main airway branches. They grow to large sizes and form cavities in the lung, but may not spread as rapidly as other lung cancers. When squamous cell cancer metastasized, it may spread to the bone, adrenal glands, liver, small intestine and brain. SCC is always caused by smoking and used to be the most common cancer. Well-differentiated SCC often grows more slowly than other cancer types. It accounts for approximately 25% of all lung cancer.

Large cell carcinoma: Large cell carcinoma (LCC) cells develop in the smaller bronchi or in scarred tissue around the outer edges of the lungs. They are unidentifiable when viewed under a microscope as squamous cell cancers or adenocarcinomas. These large cell carcinoma cells divide and replicate quickly, forming tumors that aggressively spread from the lungs to other parts of the body. It is often poorly differentiated and tends to metastasize early. It accounts for approximately 10% of all subtypes of lung cancer.

1.2.1.2 Small cell lung cancer (SCLC)

Small cell lung cancer is given this name because the cancer cells are small cells that are mostly filled with the nucleus. It was formerly referred to as "oat-cell" carcinoma. It has long been dichotomously staged into limited and extensive stage disease and responds well to chemotherapy and radiation therapy. This type of cancer is strongly caused by smoking. It accounts for approximately 15% of lung cancer. This

type of cancer is strongly caused by smoking. It is very rare for someone who has never smoked to develop it.

1.3 Lung cancer staging

There are two systems for staging lung cancer. NSCLC is categorized using the tumor, node, and metastasis (TNM) staging system [23] as summarized in **Table 1.1.** SCLC is categorized by its propensity for early metastasis and a rapid doubling time. Rather than TNM staging, a more practical scheme divides SCLC into limited and extensive disease [24].

Table 1.1 Staging classifications for lung cancer

Staging	Description
Non–small cell lung cancer	
Local	
IA (T1N0M0)	T1 tumor: ≤3 cm, surrounded by lung or pleura; no tumor more proximal than lobe bronchus
IB (T2N0M0)	T2 tumor: >3 cm, involving main bronchus ≥2 cm distal to carina, invading pleura; atelectasis or pneumonitis extending to hilum but not entire lung
IIA (T1N1M0)	N1: involvement of ipsilateral peribronchial or hilar nodes and intrapulmonary nodes by direct extension
Locally advanced	
IIB (T2N1M0, T3N0M0)	T3 tumor: invasion of chest wall, diaphragm, mediastinal pleura, pericardium main bronchus <2 cm distal to carina; atelectasis or pneumonitis of entire lung
IIIA (T1N2M0, T2N2M0, T3N1M0, T3N2M0)	N2: involvement of ipsilateral mediastinal or subcarinal nodes
IIIB (T1-4N3M0)	N3: involvement of contralateral (lung) nodes or any supraclavicular node
Advanced	
IIIB (T4N1-3M0)	T4 tumor: invasion of mediastinum, heart, great vessels, trachea, esophagus, vertebral body, carina; separate tumor nodules; malignant pleural effusion
IV (T1-4N1-3M1)	M1: Distant metastasis
Small-cell lung cancer	
Limited disease	Evidence of tumor confined to ipsilateral hemithorax; can be compassed by a single radiation port
Extensive disease	All other diseases, including metastatic disease

TNM= tumor-nodes-metastasis; Data adapted from Spira A and Ettinger D.S. [25]

1.4 GM2 activator protein (GM2AP)

1.4.1 Biosynthesis of GM2 activator protein

GM2AP is synthesized as a prepro-polypeptide. The pre-sequence or signal peptide directs the synthesis of the protein to be carried out on the rough endoplasmic reticulum (ER) with the nascent polypeptide being extruded into the lumen. The prepro form of the GM2AP is predicted to be 193 residues with a M_r of 20,808 **(Figure 1.2)**. However, it is believed that the signal peptide is cleaved from the nascent polypeptide co-translationally. For the GM2AP the signal peptide is predicted to be contained in the first 23 residues, resulting in a pro-polypeptide of 170 residues and a M_r of 18,463. In the lysosome further processing of the N terminus results in a mature chain of 162 residues with a M_r of 17,531. The GM2AP sequence also contains a single site for *N*-linked glycosylation, Asn^{63}-Val-Thr and eight Cys residues, because of the oxidizing environment of the ER form four disulfide bonds [26-29].

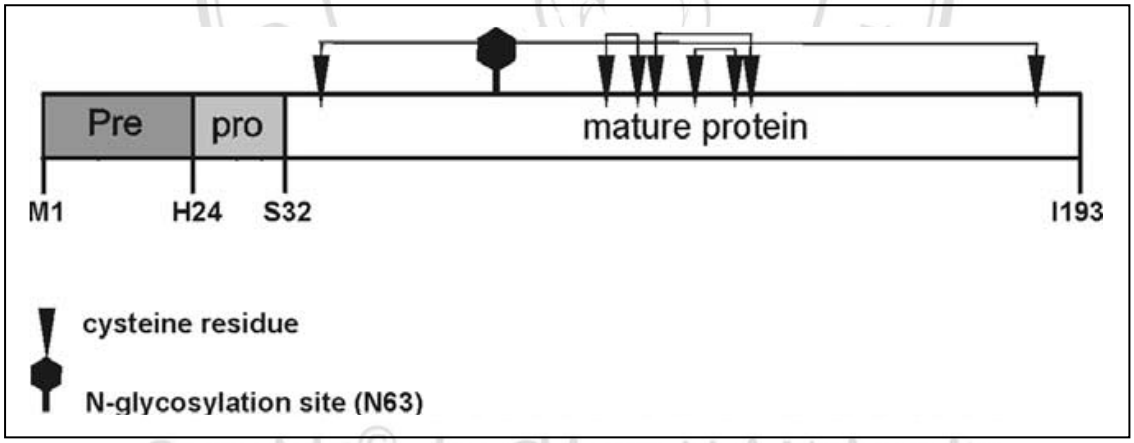

Figure 1.2 Schematic representation of the GM2AP structure. The single glycosylation site and the four intramolecular disulfide bonds are indicated. In mammalian cells, the protein is first synthesized as a 193 amino acid precursor, which is modified along the transport pathway to the lysosome. In the ER, the translocation sequence is cleaved off after Ala-23 and the pro-sequence (His-24–Leu-31) is removed after arrival in the lysosome.

1.4.2 Biological role of GM2 activator protein

The GM2 activator protein is a glycoprotein with a molecular mass of 17.6 kDa in its deglycosylated form [30]. It is acting as a cofactor essential that contains at least three functional features including a hydrophobic pocket called the β-cup structure, an oligosaccharide binding site, and an area that interacts with β-hexosaminidase A for the *in vivo* degradation of ganglioside GM2 to GM3 (**Figure 1.3**) [31, 32].

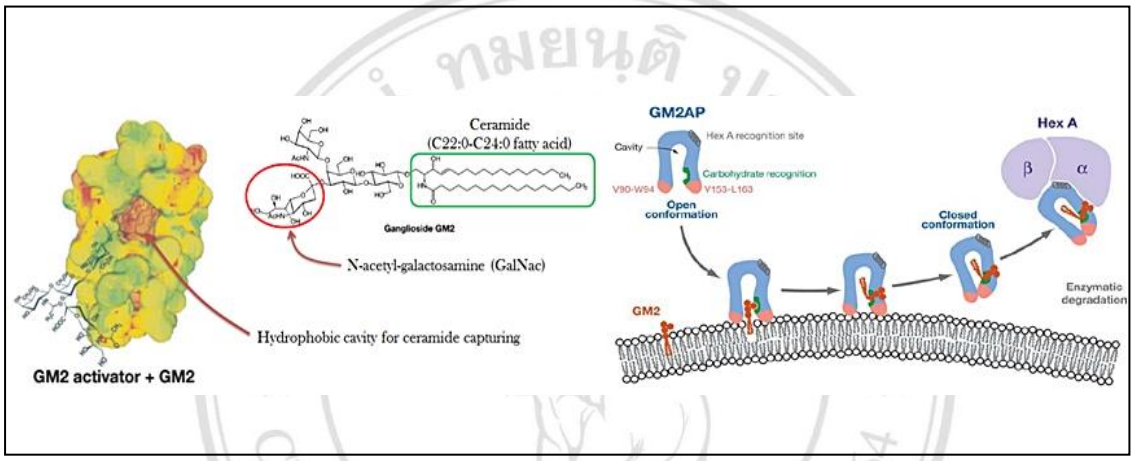

Figure 1.3 GM2AP stimulated hydrolysis of ganglioside GM2 by β-hexosaminidase A. The GM2AP contains a hydrophobic cavity, with dimensions that can accommodate the ceramide portion of GM2 and other lipids, lined by surface loops and a single short helix. The most flexible of the loops contains the substrate-binding site (V153–L163) and controls the entrance to the cavity, so that two conformations are possible: one opened and one closed. The open empty activator binds to the membrane using the hydrophobic loops and penetrates into the hydrophobic region of the bilayer. Then the lipid recognition site of the activator can interact with the substrate, and its ceramide portion can move inside the hydrophobic cavity. At this point, the conformation of the lipid-loaded activator may change to the closed one, thus the complex becomes more water soluble and leaves the membrane, exposing GM2 to the water-soluble enzyme to be degraded. β-hexosaminidases are dimeric isoenzymes formed by combination of two subunits, α and β, which differ in their substrate specificities [33]. β-hexosaminidase A can cleave glycolipid substrates on membrane surfaces only if they extend far enough

into the aqueous phase. Therefore, the degradation of ganglioside GM2 occurs only in the presence of the GM2AP.

1.4.3 GM2 activator protein and diseases

Attention in the GM2AP stemmed from its role as a substrate specific cofactor for β-hexosaminidase A. The inherited deficiency of the GM2AP leads to the AB variant of GM2 gangliosidoses (**Figure 1.4**), in which lipid accumulation in neuronal cells leads to the early death of the patients [34, 35].

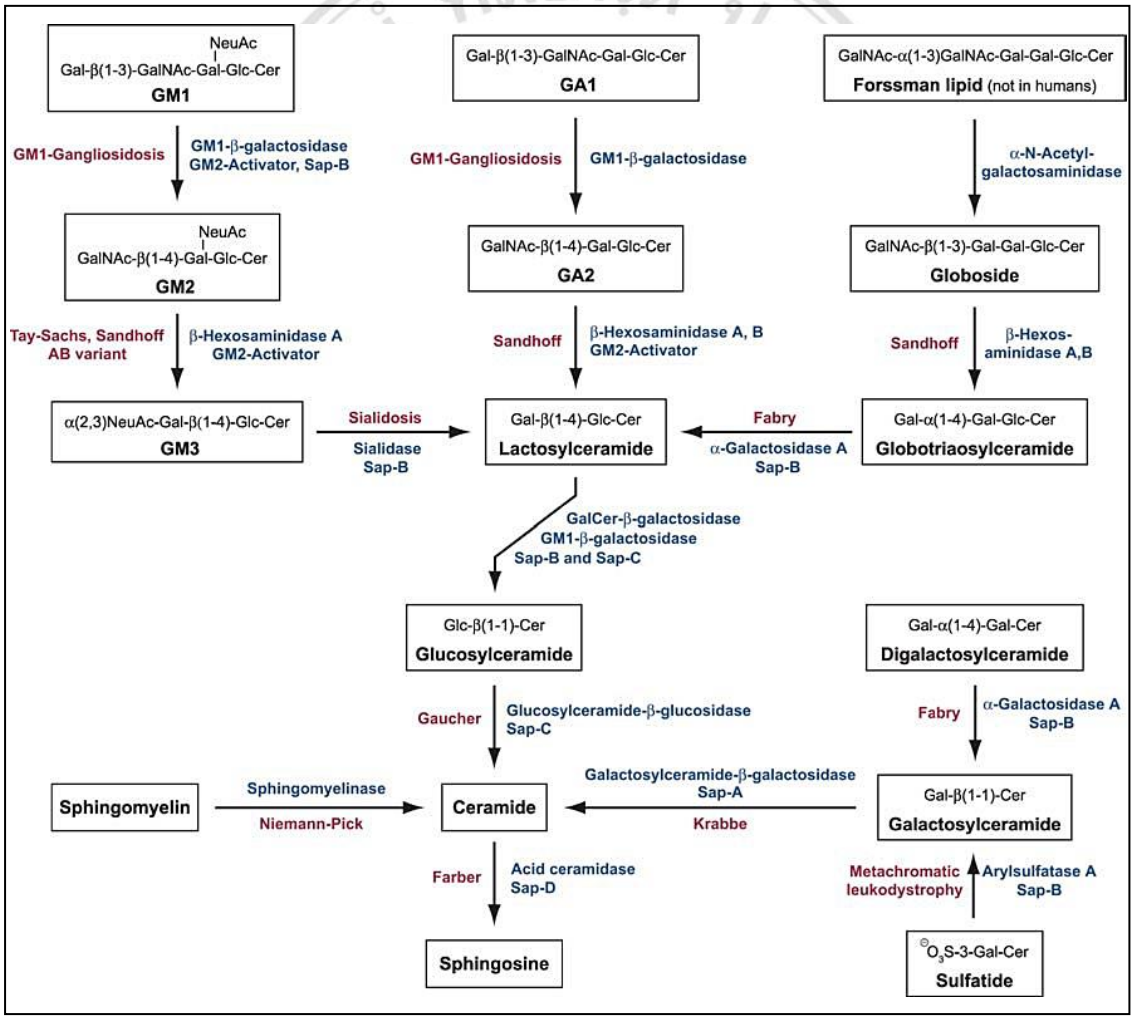

Figure 1.4 Degradation of selected sphingolipids in the lysosomes of the cells. The eponyms of individual inherited diseases (shown in red) are given. Activator proteins required for the respective degradation step *in vivo* are indicated. Variant AB: AB variant of GM2 gangliosidosis.

However, GM2AP also associated with the changing levels of ganglioside and tumor-associated gangliosides found in cancer progression. The function roles of gangliosides in tumor progression are as follows: (i) ganglioside antigens on the cell surface, or shed from the cells, act as immunosuppressors, as typically observed for the suppression of cytotoxic T cells and dendritic cells, (ii) certain gangliosides, promote tumor-associated angiogenesis, (iii) gangliosides strongly regulate cell adhesion or motility and initiate tumor metastasis, (iv) ganglioside antigens are directly connected with transducer molecules in microdomains to initiate adhesion coupled with signaling, and (v) ganglioside antigens and their catabolites are modulators of signal transduction through interaction with tyrosine kinases associated with growth factor receptors or other protein kinases [36].

1.5 Proteomics

Proteomics analysis is a powerful method for studying complex protein mixtures in microorganisms, cells, tissues, body fluids and other biological samples. Proteomics is a study of the proteome, the proteins expressed by the genome. The differential protein expression analysis can be used for protein markers potentially indicative of a disease. Currently, the combinations of 2-DE and mass spectrometry technique have become the powerful tools to find potential new protein biomarkers. Particularly, two techniques for the ionization of proteins and peptides, such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) combined with time-of-flight (TOF), as well as new hybrid mass spectrometers, that play an important role in the analysis of tumor markers [37, 38] and used for detection, identification and characterization of proteins for diagnostic biomarkers and therapeutic targets for cancer as shown in **Table 1.2**.

Table 1.2 Analytical tools in proteomics [39]

Method		Principle	Advantages	Disadvantages
Electrophoresis				
Electrophoresis		When an electric field is applied to a solution containing a protein that has a net positive or negative charge, the protein migrates at a rate that depends on its net charge, size, and shape.		Gels must be stained before proteins can be visualized. Rarely useful by itself as proteins cannot be accurately identified without the use of another detection technique such as immunoblotting or mass spectroscopy.
	SDS-PAGE	Proteins migrate through inert matrix gel of polyacrylamide. Pore size is adjustable to retard protein of interest. SDS is a negatively charged detergent that unfolds proteins and frees them from other molecules. Proteins migrate at different rates toward positive electrode.	Separates all types of proteins, even those insoluble in water.	One-dimensional separation method has limited resolution. Closely spaced bands or peaks tend to overlap. Can only resolve a small number of proteins.
	Two-dimensional gel electrophoresis	Combines 2 separation procedures. First dimension: the solubilized, denatured proteins are separated by their isoelectric point (pH where net charge is 0) in a polyacrylamide gel. Second dimension: the narrow gel containing proteins separated by isoelectric focusing undergoes electrophoresis at a right angle in SDS-PAGE to separate by size.	Good resolution of mixture. Comparison of multiple gels facilitated by image analysis software. Posttranslational modifications can be discerned. Resolution of protein approximately 1 ng/mL.	Presence of high abundance proteins (i.e., albumin, immunoglobulins) may obscure low abundance proteins. Low throughput. Final identification of protein requires spot removal from gel, digestion, and analysis of peptides by mass spectrometry. Unable to resolve low molecular weight proteins (<10,000 Da). Not easily amenable to multivariate analysis.
	Two-dimensional fluorescence difference gel electrophoresis	Labels complex mixtures with fluorescent dyes before conventional two-dimensional electrophoresis. Different cyanine dyes are used to label protein from different samples and will be excited and emit at different light wavelengths. Up to three different samples can be labeled and mixed together (test, control, reference).	Analysis of differences between mixtures is simplified. Ratio of protein expression can be obtained in a single gel, and an internal standard can be used in each gel to reduce gel-to-gel variation. Very sensitive.	Presence of high abundance proteins (i.e., albumin, immunoglobulins) may obscure low abundance proteins. Low throughput. Final identification of protein requires spot removal from gel, digestion, and analysis of peptides by mass spectrometry. Many spots cannot be identified because of lack of material. Unable to resolve low molecular weight proteins (<10,000 Da).
Protein array				
Protein arrays	Multiplex protein arrays, cytokine arrays, tissue microarrays	In most common form, antibodies to known proteins are tethered to a surface (beads, nitrocellulose, etc.) and then detected using principles of immunoassays.	High sensitivity and throughput. Multiple analytes can be measured simultaneously. Identification of potential targets already known.	Limited antibody availability and specificity. Required some prior knowledge of expressed proteins. May not detect isoforms of analyte. Cost per sample may be prohibitive.
Mass spectroscopy (MS)				
MS	Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS)	Determines the precise mass of protein or peptide fragment from protein. Protein/peptide samples are mixed with organic acid matrix, dried on metal slide, and blasted by laser ionizing the peptide, which is accelerated in an electric field toward a detector. The time it takes to reach the detector is determined by the charge and mass. Peptide sequence information can be obtained with tandem mass spectrometers (MS–MS).		Highest resolution is for molecules <3,000 Da in size.
	Surface-enhanced laser desorption/ ionization time-of-flight mass spectrometry (SELDI-TOF-MS)	Comparable to MALDI, the difference being that SELDI uses chromatographic chip arrays to selectively bind subsets of proteins from complex samples. The surfaces can be washed to remove nonspecifically bound proteins and substances that can interfere with the ionization process (salt, detergents, etc.).	High throughput via automation. Requires minimal sample preparation. Can be combined with prefractionation of material to enhance the detection of lower abundant proteins.	No direct identification of proteins. Less sensitive to high molecular weight protein (>20 kDa). May have instrument-to-instrument variation.

1.5.1 Two-dimensional electrophoresis (2-DE)

Two-dimensional electrophoresis (2-DE) is one of the most powerful and common tools for separation and fractionation of complex protein mixture extracted from tissues, cells, and other biological specimens. The principal of protein separation in 2-DE is performed into two major steps; first dimension and second dimension. In the first dimension, proteins are separated on the basis of their charge until they reach a stationary position where their net charge is zero. The pH at which a protein has zero net charge is called its isoelectric point (*pI).* Using different approaches either immobilized gradient electrophoresis (IPEG) and isoelectric focusing (IEF). In the second dimension, proteins are separated according to their molecular weight using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), the surfactant SDS binds to proteins, overriding their intrinsic charge, such that the proteins have the sample charge density and free solution electrophoretic mobility (**Figure 1.3**) [40].

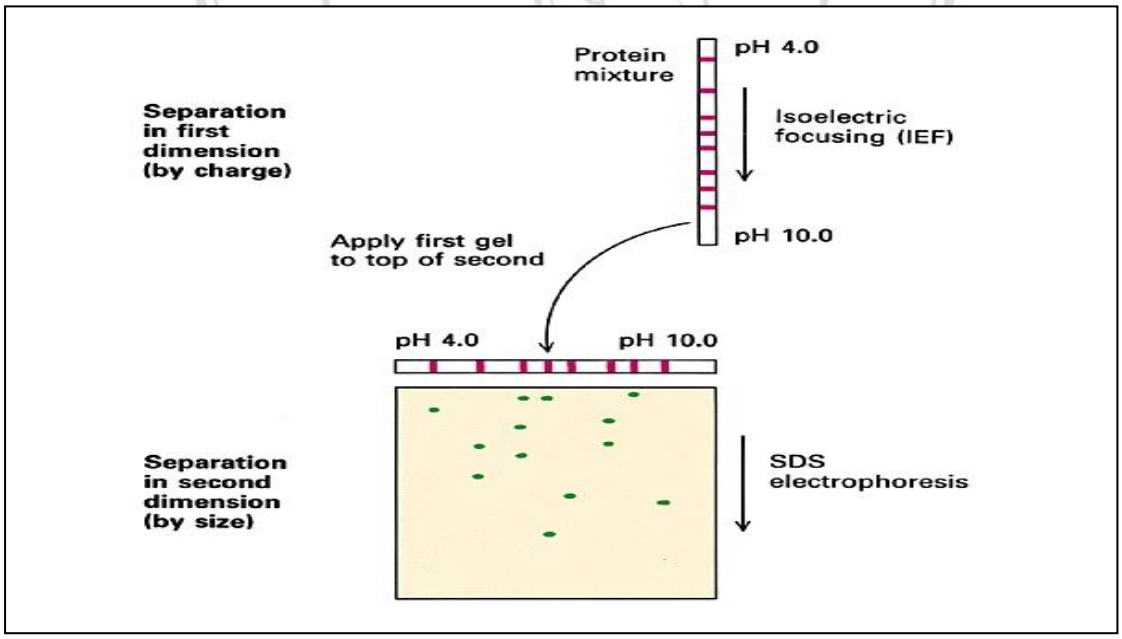

Figure 1.5 Two critical steps of two-dimensional electrophoresis (2-DE). In the first dimension, proteins are resolved in according to their isoelectric points (*pI*) using immobilized pH gradient electrophoresis (IPGE) and isoelectric focusing (IEF). In the second dimension, proteins are separated according to their approximate molecular weight using sodium dodecyl sulfate polyacrylamide-electrophoresis (SDS-PAGE).

1.5.2 Mass spectrometry (MS)

Mass spectrometry has become an important tool for protein identification, peptide sequencing, identification and location of posttranslational modifications of proteins. Two ionization techniques are commonly used for bimolecular, such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI).

1.5.2.1 Electrospray ionization mass spectrometry (ESI/MS)

ESI is a technique used electrical energy to transfer the ions from solution into the gaseous phase before they are subjected to mass spectrometric analysis. The transfer of ionic from solution into the gas phase by ESI involves three steps are dispersal of a fine spray of charge droplets, solvent evaporation and ion ejection from the highly charged droplets as shown in **Figure 1.4**. It is maintained at a high Voltage (e.g. 2.5 – 6.0 kV) relative to the wall of the surrounding chamber. The mist of highly charged droplets with the same polarity as the capillary voltage is generated [41].

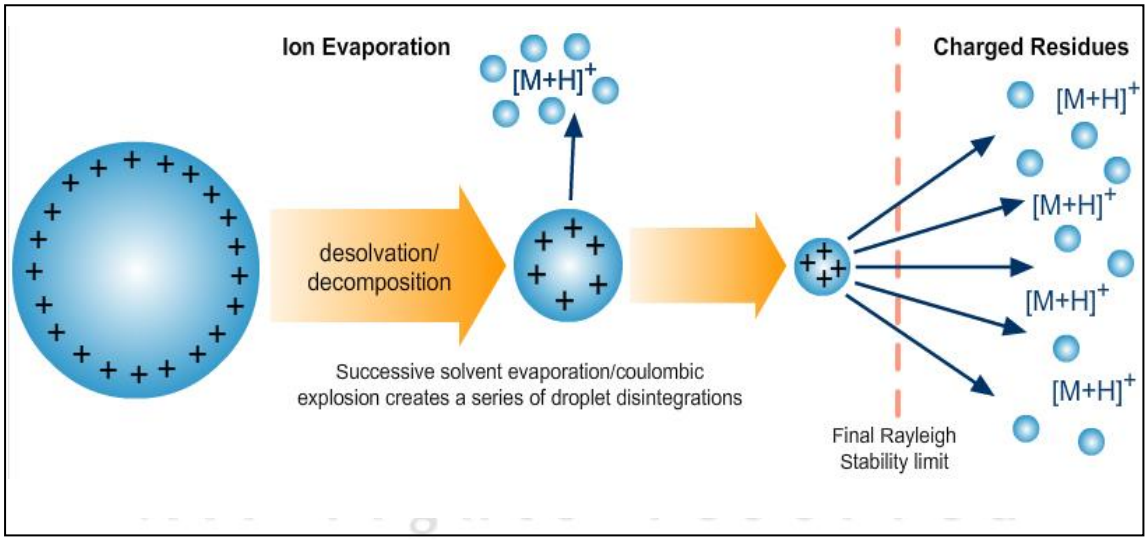

Figure 1.6 Mechanism of electrospray ionization. Within an ESI source, a continuous stream of sample solution is passed through a stainless steel or quartz silica capillary.

1.5.2.2 Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS)

MALDI-TOF MS is a technique that co-precipitates of an UV-light absorbing matrix and a bimolecular are irradiated by a nanosecond laser pulse. This technique involves spotting small concentrated aliquots of sample on to a matrix-coated target. Then, the target is positioned inside the mass spectrometer and the biomolecule of interest is desorped from the matrix surface and ionised by the laser. Most of the laser energy is absorbed by the matrix, which prevents unwanted fragmentation of the biomolecule. The ionized biomolecules are accelerated in an electric field and enter the flight tube of a time of-flight mass spectrometer. The different molecules are separated according to their mass to charge ratios and reach the detector at different times as shown in **Figure 1.5.** In this way each molecule yields a distinct signal. The method is used for detection and characterization of biomolecules, such as proteins, peptides and oligosaccharides with molecular mass between 400 and 350,000 Da. It is a very sensitive method, which allows the detection of low (10^{-15} to 10^{-18} mole) quantities of sample with an accuracy of 0.1-0.01 %. However, this technique can be very sensitive. The concentrated samples achieve the best results [41-44].

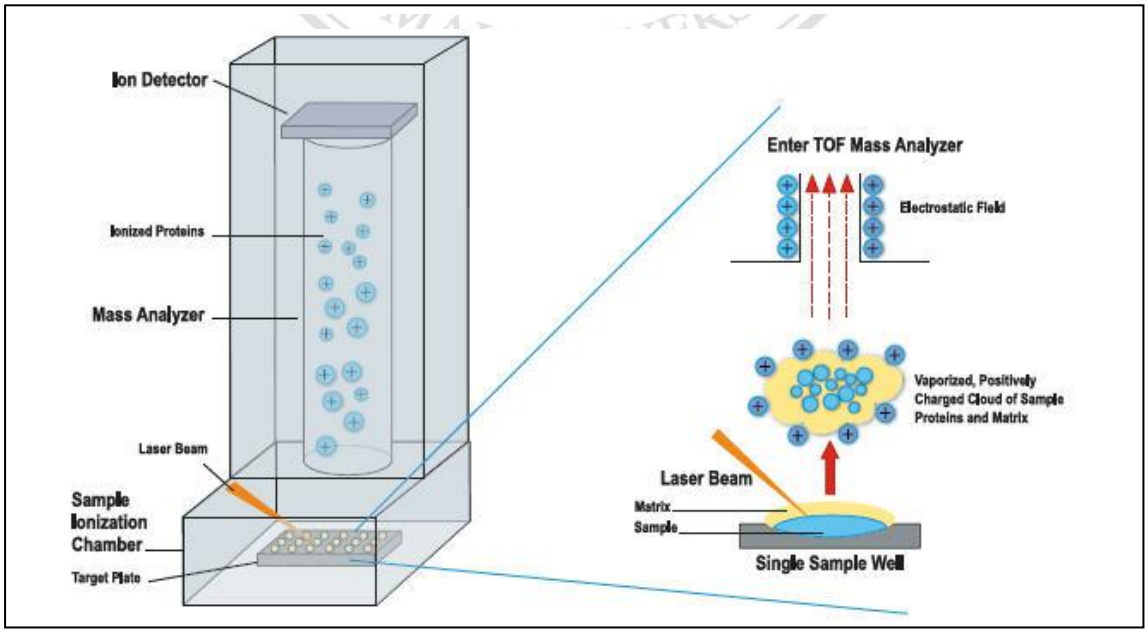

Figure 1.7 The schematic representation of a MALDI-TOF/MS

1.6 Post translational modifications

Posttranslational modifications (PTMs) are the chemical modification of a protein after its translation [45]. Protein has been thought as a linear polypeptide decorated with complicated modifications. PTMs affect the physicochemical properties of proteins which provide a mechanism for the dynamic regulation of molecular self-assembly and catalytic processes through the reversible molecular recognition of proteins, nucleic acids, metabolites, carbohydrates and phospholipids as shown in **Figure 1.5** [9, 46, 47].

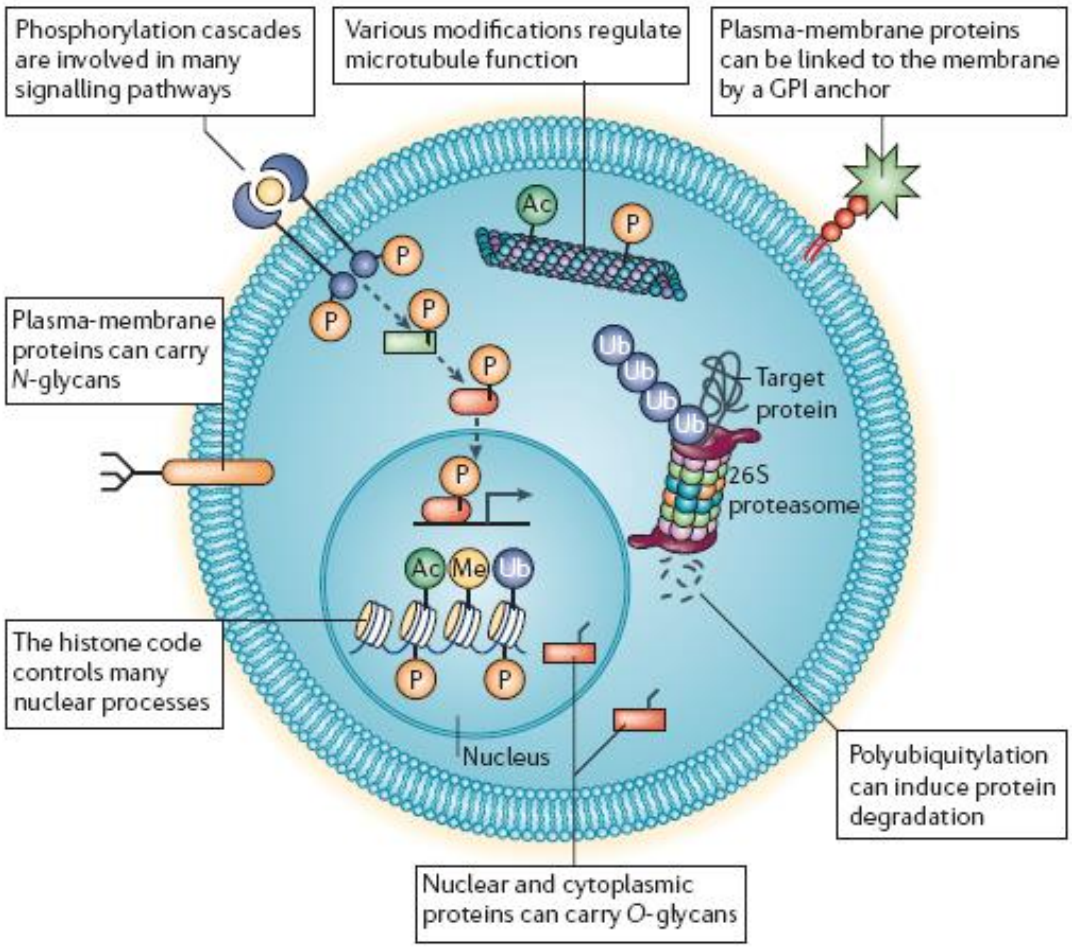

Figure 1.8 Cellular posttranslational modifications. This schematic shows the location and role of a selection of some of the most important of PTMs.

Protein glycosylation is one of the most common PTMs in current cancer research and plays a fundamental role in a diversity of biological processes. Importantly, it correlates with tumor development and malignancy [9, 48–50]. More than 50% of proteins in human are glycosylated [51]. The altered glycosylation is presented on the surface of the cell as a part of the glycoproteins and glycolipids and is shaded into the bloodstream, urine, rectal secretion, or saliva as soluble tumor glycoproteins [52, 53]. Changes in glycosylation often arise from changes in the expression level of glycosyltransferase in the Golgi apparatus of cancer cells, resulting modifications in the core and terminal structure of *N*-linked and *O*-linked glycans [54]. Glycan chains in glycoproteins play key function in biological processes and have influence on the physicochemical properties such as folding, stability and propensity of these proteins to be degraded by protease activity can be affected by the presence or absence of carbohydrate chains (**Table 1.3**). Therefore, change in abundance and alterations in glycan profiles of body fluids or cell surface proteins have been correlated with progression of cancer and other disease states [55, 56].

Table 1.3 The potential role and effect of the glycocomponent of glycoproteins [57]

Role/effect	Comment
Protein folding	Glycosylation can affect local protein secondary structure and help direct folding of the polypeptide chain.
Protein targeting/trafficking	The glycocomponent can participate in the sorting/directing of a protein to its final destination.
Ligand recognition/binding	The carbohydrate content of antibodies, for example, function and antibody binding to monocyte Fc receptors and interaction with complement component Cl_q.
Biological activity	The carbohydrate side chain of gonadotrophins (specifically the α-subunit N^{52} side chain) is essential to the activation of gonadotrophin signal transduction.
Stability	Sugar side chains can potentially stabilize a glycoprotein in number f ways including enhancing its solubility, shielding hydrophobic patches on its surface, protecting from proteolysis and directing participation in intrachain stabilizing interactions.
Regulates protein haft-life	Large amounts of sialic acid can increase a glycoprotein's plasma haft-life. Exposure of galactose residues can decrease plasma haft-life by promoting uptake through hepatic galactose residues. Yeast N-glycosylation is of a high-mannose type, driving rapid removal from circulation through mannose receptors.
Immunogenicity	Some glycosylation motifs characteristic of plant-derived glycoproteins (often containing fucose and xylose residues) are highly immunogenic in mammals.

1.7 Purpose of the current study

This study aims to differentiate protein expression in healthy controls and lung cancer patients that focusing on GM2AP. The first purpose is to compare the protein profiles between healthy and lung cancer urine samples. The second purpose is to measure the expression of GM2AP level in urine, serum and tissue specimens of lung cancer patients, and to correlate GM2AP expression with clinicopathological features of lung cancer for protein marker**.** The third purpose is to investigate the glycan structures of urinary GM2AP in lung cancer patients.