IDENTIFICATION OF BIOMARKERS IN INFILTRATING DUCTAL CARCINOMA: A PROTEOMICS APPROACH

by

SENG LIANG

Thesis submitted in fulfillment of the

requirements for the degree of

Master of Science

January 2009

ACKNOWLEDGMENT

I would like to thank some people who have helped a lot me in this study. First of all, I would like to express my deepest gratitude to my supervisor, Dr Gam Lay Harn for her guidance, help and knowledge in completing this project. I would also like thank all the lecturers for lending their experiences and offering their insights and advices during the course of the study.

Secondly, I would like to thank the staffs of the School of Pharmaceutical Sciences, namely En. Basri Jaafar, En. Rosli Kassim, En. Tan Siao Peng, En. Fizal Jamaludin, Pn Salida, En. Annuar Apendi Ahmad and En. Hassan for kindly providing me with assistance, facilities and equipment in my experimental work.

I would like to thank The Institute of Graduate Studies for providing me with the Graduate Assistant Fund and the Fellowship Scholarship Scheme.

I would also like to extend my gratitude to all my labmates for being there when I need them. Thank you very much for your help and support. My special thanks to Dr Saravanan Dharmaraj for providing assistance with the statistical analysis of the results.

I would like to express my gratitude as well to the Penang General Hospital and Mr Manji Singh for kindly providing breast tissue samples for my analysis. I would also like to acknowledge the National Poison Centre for providing me with the facilities to analyze the results of the project.

Finally, to my family, a big thank you for always being there for me during the course of the study, especially to my mother for her love, understanding and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACN	: Acetonitrile
AEBSF	: 4-(2-Aminoethyl) benzenesulfonyl fluoride
APS	: Ammonium persulfate
BPC	: Base peak chromatogram
BSA	: Bovine serum albumin
CHAPS	: 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CID	: Collision induced dissociation
4 – CN	: 4 – Chloronaphthol
DTT	: 1,4 – Dithiothreitol
ESI	: Electrospray ionization
HPLC	: High performance liquid chromatography
IAM	: Iodoacetamide
IDC	: Infiltrating ductal carcinoma
IEF	: Isoelectric focusing
IPG	: Immobilized pH Gradient
kDa	: kilo Dalton
LDA	: Linear discriminant analysis
LC-MS/MS	: Liquid chromatography tandem mass spectrometry
MS/MS	: Tandem mass spectrometry
MS	: Mass spectrometry
m/z	: Mass to charge
$Na_2HPO_4.2H_20$	: Disodium hydrogen phosphate dihydrate
$NaH_2PO4.2H_20$	: Sodium dihydrogen phosphate dihydrate

PAGE	: Polyacrylamide gel electrophoresis
PCA	: Principal component analysis
pI	: Isoelectric point
RT	: Room temperature (25°C)
RP	: Reverse-phase
SDS	: Sodium dodecyl sulfate
SPSS	: Statistical Package for Social Science
TEMED	: N, N, N’ N’ – tetramethylethylenediamine
TLB	: Thiourea lysis buffer
TRIS	: Tris buffer
Tris	: Tris(hydroxymethyl)aminomethane

LIST OF APPENDICES

PENGENALPASTIAN PENANDA-PENANDA BIO DALAM KARSINOMA DUKTAL MENYUSUP: SATU PENDEKATAN PROTEOMIK

ABSTRAK

Kanser ialah suatu penyakit disifatkan dengan pertumbuhan sel-sel yang tidak terancang. Payudara wanita mengandungi kelenjar mamari yang dikhaskan untuk merembes susu semasa wanita mengandung. Kanser payudara wanita merupakan salah satu punca utama kematian wanita di seluruh dunia. Di Malaysia, kanser payudara merupakan kanser yang paling lazim didiagnosiskan. Dalam hal ini, kaum Cina merangkumi kes kanser payudara yang paling banyak, diikuti dengan kaum India dan kaum Melayu. Jenis kanser payudara yang paling lazim ialah karsinoma duktal menyusup (IDC). Dalam kajian ini, satu pendekatan proteomik dilakukan untuk menganalisisa protein-protein dari tisu-tisu payudara manusia. Teknik pengektrakan protein secara bersambung digunakan untuk mengekstrak protein daripada tisu-tisu dan seterusnya protein-protein ini dipisahkan dengan elektroforesis gel poliakrilamida dua-dimensi (2D-PAGE). Dalam kaedah ini, protein-protein dipisahkan mengikut takat isoelektrik (pI) dan jisim molekul masing-masing. Gel-gel diwarnakan dengan pewarna biru Coomassie dan imej-imej gel diambil dan analisa imej-imej dijalankan dengan menggunakan sistem pengimej. Keamatan tompok-tompok protein sasaran dianalisa dengan menggunakan analisis komponen-komponen utama (PCA) dan analisis pembezalayan linear (LDA) untuk menghitung kesahihan protein-protein ini sebagai penunjuk dalam IDC. Tompok-tompok protein sasaran dipotong daripada gel dan diproses dengan menggunakan teknik penghadaman dalam-gel dengan enzim tripsin. Peptida-peptida hasil hadaman

tripsin dianalisisa dengan LC-MS/MS. Spektrum ion produk yang dihasilkan dalam imbasan MS/MS dicarikan pada pangkalan carian data protein MASCOT enjin carian. Dua puluh empat protein yang larut air dan sepuluh protein hidrofobik didapati diekspres dalam kadar yang berbeza antara tisu-tisu payudara normal dan kanser. Antara peranan protein yang dikenalpastikan adalah dalam pertumbuhan dan penyelenggaran sel, komunikasi sel, transduksi isyarat, metabolisma dan pengangkutan. Oleh sebab protein ialah komponen berfungsi yang mengawalatur aktiviti sel, pengenalpastian protein-protein ini memberi pemahaman yang lebih mendalam tentang perkembangan kanser payudara pada pesakit-pesakit Malaysia.

IDENTIFICATION OF BIOMARKERS IN INFILTRATING DUCTAL CARCINOMA: A PROTEOMICS APPROACH

ABSTRACT

Cancer is a disease characterized by disordered cell growth. The female breast houses the mammary glands which are specialized to secrete milk following pregnancy. Female breast cancer is one of the leading causes of female mortality worldwide. In Malaysia, breast cancer is the most commonly diagnosed cancer in women. Of these, the Chinese had the most number of breast cancer cases, followed by the Indians and the Malays. The most common type of breast cancer is infiltrating ductal carcinoma (IDC). In this study, a proteomics approach was used in the analysis of proteins from human breast tissues. Sequential protein extraction technique was used to extract the proteins from the tissues and the proteins were separated by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). Using this method, the proteins were separated according to their isoelectric point (pI) and molecular mass. The gels then were stained using Coomassie Blue and their images were captured and analyzed by an imaging system. The target proteins spots' intensities were analyzed using principal components analysis (PCA) and linear discriminant analysis (LDA) for their significance in indicating IDC. The targeted protein spots were excised from the gel and were subjected to in-gel digestion using the enzyme trypsin. The tryptic peptides were analyzed by LC/MS/MS. The generated product ion spectrum in the MS/MS scan was searched against the MASCOT protein database search engine. Twenty-four aqueous soluble proteins and ten hydrophobic proteins were found to be differentially expressed in normal

and cancer breast tissues. Among the roles the identified proteins were in cell growth and maintenance, cell communication, signal transduction, metabolism and transport. As proteins are the functional components that regulate cellular activity, identification of these proteins provides a better understanding to the development of breast cancer amongst Malaysian patients.

CHAPTER 1

INTRODUCTION

1.1 Cancer

1.1.1 Definition

Cancer refers to diseases characterized by a common type of disordered cell growth (Creasey, 1981; King and Robins, 2006). This normally leads to invasion and destruction of the surrounding tissues or spread to other parts of the body (King and Robins, 2006). The disease is usually visible as a mass of cells called tumour as a result of a series of biological and chemical changes which may take years to develop (Franks, 1986). Cancer may affect people regardless of age and sex. Out of 61 million cases of death reported worldwide in 2007, 7.9 million or 13% is caused by cancer, making it one of the world's leading causes of death (World Health Organization, 2008).

Tumours, strictly defined as neoplasms, are classified as benign, in *situ* and malignant. Tumours are more commonly exhibited in epithelial cells (King and Robins, 2006). Benign tumours are non-cancerous and are usually harmless. They are composed of well-differentiated cells (Vincent, 1985) and do not infiltrate adjacent tissues and remain non-invasive (Creasey, 1981). *In situ* tumours usually develop in the epithelium and are usually small. They have the morphological appearances of cancer cells but remain in the epithelial layer. They do not invade the basement membrane and the supporting mesenchyme. However, malignant tumours are life threatening as they are capable of invading and destroying nearby tissues. They also display cellular abnormalities (Franks, 1986). A group of cells in a

malignant tumour may also break away and spread to other parts of the body, a process called metastasis (Hart, 1986; Ruddon, 1995; King and Robins, 2006).

Many different factors are involved in the potential development of cancer. They include: age, race, family history, personal history of cancer, presence of viruses, mutations in cell regulation genes and tumour suppressing genes, exposure to carcinogens, lack of physical activity and diet (Franks, 1986; National Cancer Institute, 2007a).

Cancers are classified according to the type of cell it originates from and its location. The most common form of malignancy is those derived from epithelial cells and is called carcinoma. The other major type of cancer involves the mesenchyme or connective tissue cells are called sarcoma (Creasey, 1981, Jass, 1999). As of now, there is no known cure for cancer. Cancer treatments available include surgery, radiotherapy, chemotherapy, hormonal therapy and immunotherapy (Creasey, 1981; Franks, 1986; King and Robins, 2006).

1.1.2 Carcinogenesis

Carcinogenesis is a multistage process of cancer development from a normal single cell. Carcinogenesis can be divided into three main stages: initiation, promotion and progression (King and Robins, 2006). In the initiation stage, normal cells are exposed to cancer causing agents or carcinogens for a period of time. This stage occurs rapidly and may persist for a long time. The primary consequence of carcinogen exposure is likely to be the damage to the cells' genetic material. Initiated cells usually remain latent until acted upon by promoting agents (Franks, 1986).

As there are a series of changes after the initiation stage, tumour formation may not happen immediately. Tumour promotion may be induced by the carcinogen or by other promoting agents which by themselves do not cause tumour formation. Here, the slow growing 'transformed' cells are induced by the promoting agents to divide and grow rapidly. Many agents can cause cell division, but only promoters are able to initiate tumour development by interfering with the cell's differentiation process. Even as this happens, the cells may still be acted upon by normal growth inhibiting factors in the body so that the final outcome depends on the balance between the factors and the extent of changes within the cell (Franks, 1986).

In the progression stage, the tumours will undergo further changes such as dedifferentiation, increased autonomous growth and aggressive behaviour. This stage reflects the multiple changes in growth regulatory mechanisms and ultimately results in autonomous cell growth, which is the ability to grow outside their normal environment and spread to other parts of the body (King and Robins, 2006).

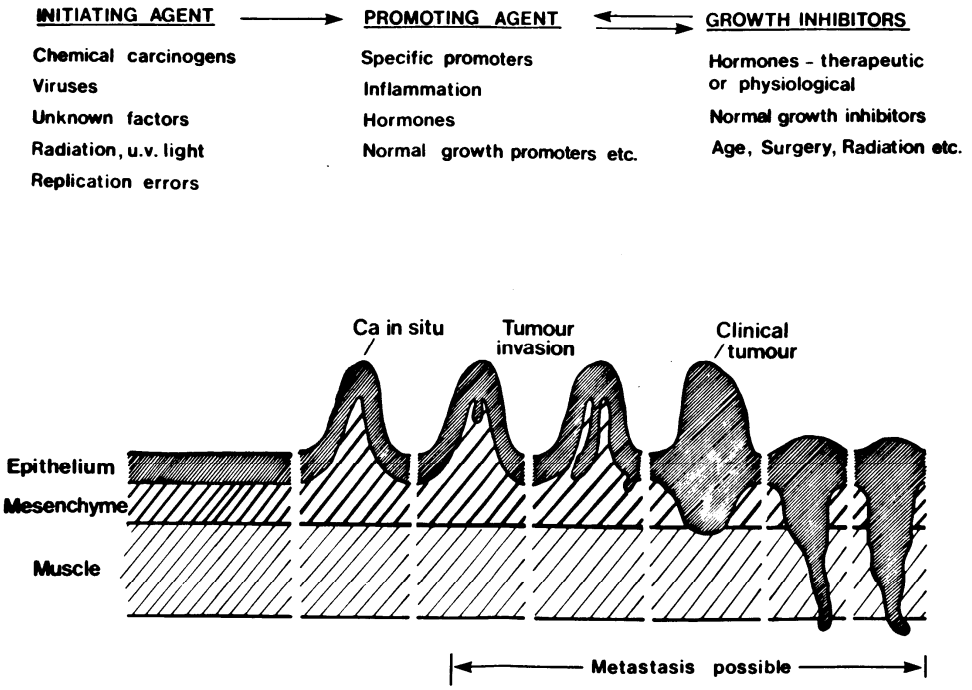

Figure 1.1 Factors influencing tumour development showing progression from normal to invasive tumour (Franks, 1986).

Normal growth control is regulated by two different genes: oncogenes, which function to accelerate cell division, and tumour suppressing genes, which does the opposite. Under normal conditions, the relationship between these two processes is delicately balanced such that cells undergo controlled patterns of growth and division. Cancer development is associated with an accumulation of mutations of these genes. The carcinogen is believed to have caused the damage or mutation of these genes which disturb this balance in favour of cell growth and proliferation (Sutherland, 1999). Besides these genes, cancers may form as a result of mutation to proto-oncogenes, cell cycle regulator genes, housekeeping genes, metastasis genes and drug resistance and susceptibility genes (Trent, 1999). In addition, defects in gene repair mechanisms may induce genome-wide instability, which may lead to further progression of cancer (Baak *et al.*, 2003).

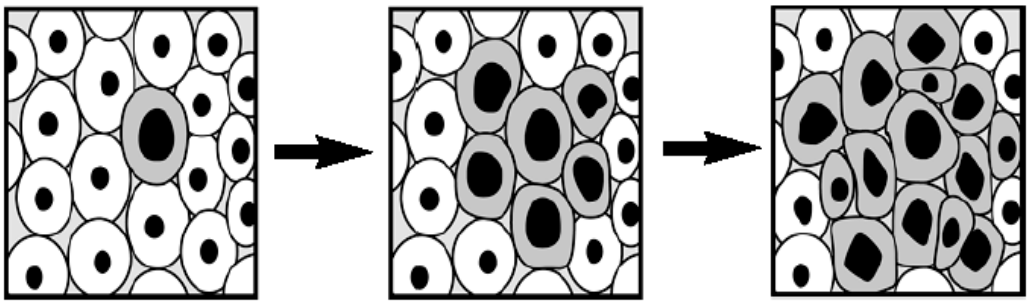

Figure 1.2 Growth of cancer. Cancer cells (shaded) grow and multiply to form tumours inside the tissue. (Susan G. Komen For The Cure, 2007).

1.1.3 Metastasis

Metastasis is a process when cancer cells escape from the primary organ or tissue where the tumour initially occurs and spread to other parts of the body (King and Robins, 2006). At the new location, cancer cells may again begin to divide abnormally to form a secondary site. Metastasis usually forms part of the natural

history of malignant tumours and is the major cause of death in cancer. The ability of malignant cells to break away and invade surrounding tissues represents the main difference between benign and malignant tumours (Gallagher, 1985; King and Robins, 2006).

Metastasis occurs in several steps. Following tumour development and growth, the cancer cells will invade and infiltrate the surrounding tissue, traverse the basal membrane and migrate through the intracellular space to reach the blood or lymphatic vessels. They penetrate the vessel walls (intravasation) in order to release the tumour cells into the circulation. Once in circulation, they have to avoid destruction while being transported to a new location. When the cells have arrived at a remote site, they arrest in the capillary beds of the remote site and then breach the blood or lymphatic vessel walls again (extravasation). Having reached the metastatic site, the tumour cells will proliferate in the new environment. If these steps are completed, the result will be a formation of a secondary tumour in a distant part of the body (Hart, 1986: King and Robins, 2006).

The critical steps in metastasis are illustrated in Figure 1.3.

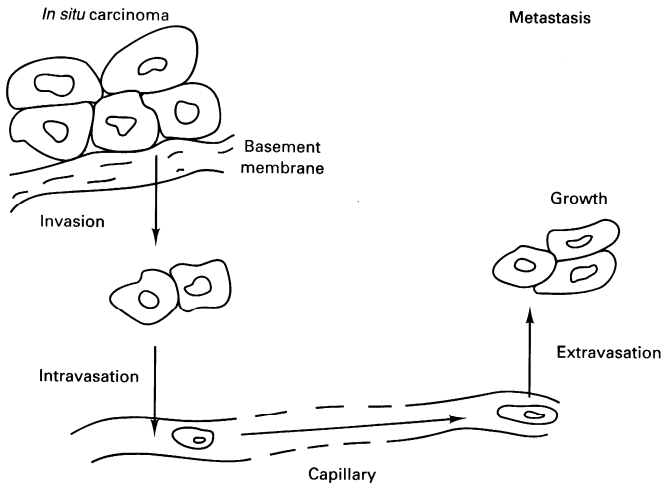

Figure 1.3 The process of metastasis (Gallagher, 1985).

There are three modes of transporting tumour cells from a primary site to a secondary site: blood vessels, lymphatic vessels and movements within the body cavities (King and Robins, 2006).

1.2 The Breast

The breast is a hemispherical-shaped elevation of the anterior thorax which houses the mammary glands, which are accessory organs of the female reproductive system specialized to secrete milk following pregnancy. They are located in the subcutaneous tissue of the anterior thorax. The breasts overlie the pectoralis major muscles and extend from the second to the sixth ribs and from the sternum (breastbone) to the axillae (underarm). A nipple is located near the tip of each breast and is surrounded by a circular pigmented area called the areola (Hole, 1990).

A mammary gland is composed of 15 to 20 irregularly shaped lobes divided by connective and adipose (fatty) tissues. Each lobe is subdivided into lobules (Giometti *et al.*, 1995). The lobules contain glandular alveoli that secrete milk during lactation. The alveoli secrete milk into a series of secondary tubules, which converge to form a series of mammary ducts. These converge to form a lactiferous duct that drains at the tip of the nipple. However, just beneath the nipple's surface, the lumen of each lactiferous duct expands to form an ampulla, where milk accumulates during nursing (Fox, 1999).

The connective tissues support the mammary glands and attach them to the fascia of the underlying pectoral muscles. Other connective tissue, called suspensory ligaments, extends inward from the skin to the fascia and helps support the breast's weight. The amount of fatty tissue only determines the shape and size of a breast (Hole, 1990; Fox, 1999). The lactiferous ducts consist of a basal and a luminal layer.

The basal layer contains myoepithelial cells and a pluripotent stem cell population from which basal, luminal and secretory epithelial cells differentiate (Dulbecco *et al.*, 1986).

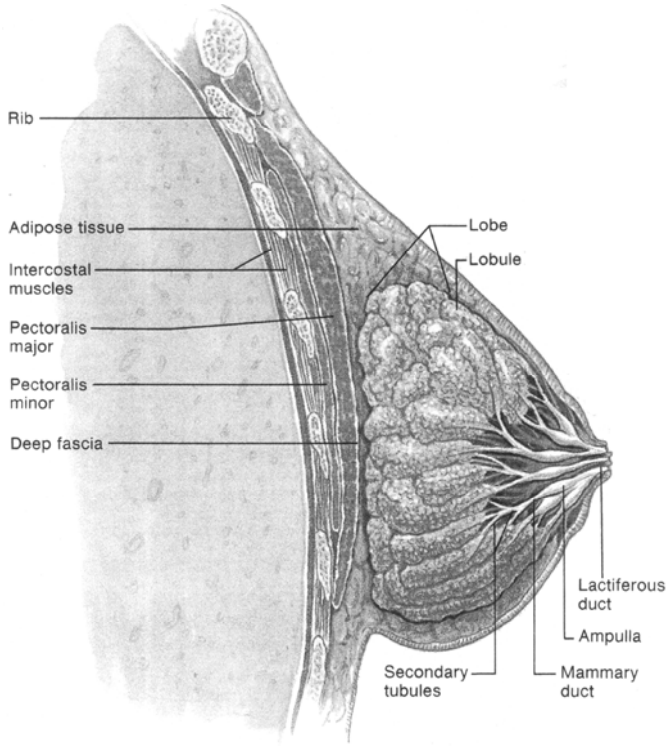

Figure 1.4 Saggital view of the structure of the breast and mammary glands (Fox, 1999).

1.3 Breast Cancer

Breast cancer is a condition where malignant tumour cells form in tissues of the breast, usually the ducts and lobules. About 85 percent of breast cancers originate in the mammary ducts, while about 15 percent arise in the lobules (Dickson, *et al.*, 2005). It occurs in both men and women, although it is rare in males (National Cancer Institute, 2008). Although these tumours maybe localized, they may metastasize to other parts of the body to form secondary tumours.

The most common symptoms in breast cancer are a change of how the breasts or nipples feel. Most commonly, there is a lump or thickening in or near the breast or in the underarm area. Other symptoms are nipple changes such as

tenderness or discharge. In some cases, changes in the size, contour or shape of the breast, a nipple turned inward and changes in the skin of the breast, areola or nipple. In advanced stages, the tumour growth may ulcerate through the skin and become infected. Bone pain, liver tenderness, severe headaches, shortness of breath and a chronic persistent cough may be an indication that the cancer has spread (Gillett, 1999; Malaysian Oncological Society, 2007; National Cancer Institute, 2007b; Oncolink, 2008).

Early stage detection of breast cancer is often by breast self-examination. Commonly used screening methods for breast cancer are mammography, magnetic resonance imaging (MRI), ultrasound and needle biopsy (Nancer Cancer Institute, 2005c; Harris, 2005; Malaysian Oncological Society, 2007). After a microscopic examination of the cancer cells, the stage and grade can be established. Common forms of treatment for breast cancer are conservative surgery, mastectomy, radiotherapy, chemotherapy, hormonal therapy and targeted therapy (Malaysian Oncological Society, 2007; National Cancer Institute, 2007d). More recent forms of advanced treatment which are less invasive include: Radiofrequency ablation, cryoablation, interstitial laser ablation and focused ultrasound ablation (Huston and Simmons, 2005).

1.3.1 Major Forms of Breast Cancer

The most common forms of breast cancer originate in the ducts or lobules of the breast. The cancer's point of origin is determined by an examination of the cells' microscopic appearance taken from a biopsy. In general, breast cancer can be categorized as *in situ* or non-invasive and infiltrating or invasive.

In *in situ* breast cancer, the cancer cells are confined within their point of origin and have not invaded the surrounding breast tissues. Two common forms of in situ breast cancer are ductal carcinoma *in situ* and lobular carcinoma *in situ*.

Ductal carcinoma *in situ* (DCIS) is a condition where the tumour cells develop in the lining of the lactiferous duct but have not yet invaded the surrounding breast tissue. Detected early, almost all DCIS cases can be successfully treated. DCIS is considered an early-stage breast cancer but may progress to infiltrating ductal carcinoma if left untreated. DCIS accounts for 90% of all in situ breast cancer cases (Gillett, 1999; MayoClinic, 2007).

Lobular carcinoma *in situ* (LCIS) is a tumour of the lobular cells in the breast which has not invaded the surrounding tissue. It does not usually progress to invasion but patients with LCIS are thought to be at an increased risk of developing invasive breast cancer at some point in their future (Gillett, 1999; MayoClinic, 2007).

Infiltrating or invasive breast cancers are capable of breaking free from their point of origin and invade the surrounding breast tissues. They also have the ability to travel to other parts of the body causing secondary tumours. Two major forms of infiltrating breast cancer are infiltrating ductal carcinoma and infiltrating lobular carcinoma.

In infiltrating ductal carcinoma (IDC), tumour cells formed in the lining of the lactiferous duct break free of the ductal wall and invade the surrounding tissue. IDC is the most common form of invasive breast cancer, accounting for about 85% of all cases (Molland *et al.*, 2004; Dickson, *et al.*, 2005; American Cancer Society, 2006a; Oncolink, 2007). The tumour cells in IDC may remain localized or they may metastasize to other parts of the body (MayoClinic, 2005)

In infiltrating lobular carcinoma (ILC), the tumour cells originate in the lobules of the breast and invade the surrounding tissue. As with IDC, they can metastasize to other parts of the body. Unlike IDC, it is harder to detect as a lump my not be detected and is likely to appear on a mammogram (MayoClinic, 2007).

1.3.2 Less Common Forms of Breast Cancer

Inflammatory breast cancer is a rare but aggressive type of breast cancer where the tumour cells block the lymph vessel near the surface of the breast resulting in reddening and swelling of the skin of the breast. Medullary carcinoma is a specific type of invasive breast cancer where the boundary between tumour and normal cells are clearly defined. Mucinous or colloid carcinoma is an invasive breast cancer where the tumour cells produce mucus and the tumour takes on a jelly-like appearance. Paget's disease is a rare type of breast cancer that affects the nipple and the areola. Tubolar carcinoma is a rare form of breast cancer. Phylloides tumours develop in the connective tissue of the breast rather than the lobules or ducts. Metaplastic carcinoma tends to remain localized and contain several different types of cells that are not usually seen in other forms of breast cancer. It represents less than 1% of all newly diagnosed breast cancers (Oncolink, 2001; MayoClinic, 2007).

1.3.3 Breast Cancer Stages

Staging is a system of classifying patients based on the extent of the cancer in the primary site and how much it has spread to other parts of the body (National Cancer Institute, 2007e). Staging defines the extent of spread of a cancer (Jass, 1999). This information is used to plan treatment and to find out a patient's prognosis. To overcome weakness of older staging systems, the American Joint

Committee on Cancer (AJCC) developed the TNM classification system as a tool for doctors to stage different types of cancer based on certain standard criteria (Greene *et al.*, 2002).

The TNM system is based on the extent of the tumour (T), the extent of spread to the lymph nodes (N), and the presence of metastasis (M). A number is added to each letter to indicate the size or extent of the tumour and the extent of spread (National Cancer Institute, 2007e).

The evaluation of cancer staging follows the T, N, M system as shown below:

Primary Tumor (T)

TX: Primary tumour cannot be evaluated

T0: No evidence of primary tumour

Tis: Carcinoma *in situ*

T1, T2, T3 or T4: Size and/or extent of the primary tumour

Regional Lymph Nodes (N)

NX: Regional lymph nodes cannot be evaluated

N0: No regional lymph node involvement

N1, N2 or N3: Involvement of regional lymph nodes

Distant Metastasis (M)

MX: Distant metastasis cannot be evaluated

M0: No distant metastasis

M1: Distant metastasis

Once the T, N, and M have been established, they are combined and an overall breast cancer stage is assigned.

Table 1.1 Breast cancer staging system (Greene *et al.*, 2002).

Stage	Description
0	Lobular or ductal carcinoma in-situ. Contained cancer which has not spread through the ductal system.
I	Tumour is smaller than 2cm and has not spread to any lymph nodes.
IIa	Tumour is smaller than 2cm, spread to 1-3 underarm lymph nodes. Tumour is between 2-5cm and has not spread to any lymph nodes.
IIb	Tumour is between 2-5cm. spread to 1-3 underarm lymph nodes. Tumour is larger than 5cm and has not spread to lymph nodes and does not grow into chest wall.
IIIa	Tumour is smaller than 5cm, spread to 4-9 underarm lymph nodes. Tumour is larger than 5cm and has spread to 1 to 9 axillary nodes or to internal mammary nodes.
IIIb	Tumour has grown into the chest wall or skin and may have spread to other lymph nodes or as many as 9 underarm nodes. Mammary nodes may or may not be involved.
IV	Tumour has spread out to other parts of the body (Metastasis).

1.3.4 Breast Cancer Grades

Grading is a way of classifying tumour cells by how abnormal they appear under a microscope and how quickly they are likely to spread (National Cancer Institute, 2004). The grading system determines the biological aggressiveness of cancer cells (Jass, 1999). Patients are assigned a grade to their breast cancer to identify the type of tumour present and help determine the patient's prognosis. The AJCC has a general guideline on cancer grading classifications. The Scarff-Bloom-Richardson system is one the most common type of breast cancer grade system used. Three features are observed when determining a cancer's grade: tubule formation (the percentage of cancer composed of tubular structures), nuclear pleomorphism (change in cell size and uniformity) and the frequency of cell mitosis

(rate of cell division). Each of these features is assigned a score ranging from 1 to 3 where 1 indicates slower cell growth and 3 a faster cell growth (American Cancer Society, 2006b; Oncolink, 2007; Imaginis, 2008).

Table 1.2 Scarff-Bloom-Richardson grading system (Oncolink, 2007; Imaginis, 2008).

Tubule Formation	Score
More than 75%	1
Between 10% and 75%	2
Less than 10%	3
Nuclear Pleomorphism	Score
Small, uniform cells	1
Moderate increase in size and variation	2
Marked variation	3
Frequency of cell mitosis	Score
Up to 7	1
Between 8 and 14	2
15 or more	3

The scores of each of the cells' features are added together for a final sum ranging from 3 to 9.

Table 1.3 Grading of breast cancer by score (American Cancer Society, 2006b; Oncolink, 2007; Imaginis, 2008).

Grade	Description	Score
1	Well-differentiated cells; cells generally appear normal and are not growing rapidly; cancer arranged in small tubules.	3,4,5
2	Moderately-differentiated cells; have characteristics between Grade 1 and Grade 3 tumours.	6,7
3	Poorly differentiated breast cells; cells do not appear normal and tend to grow and spread more aggressively.	8,9

1.3.5 Risk Factors in Breast Cancer

There is still no explanation as to the actual causes of breast cancer. Research has shown that certain risk factors, both inherited and environmental can predispose a person to develop breast cancer (National Cancer Institute, 2007a).

Women are 100 times likely to get the disease than men (American Cancer Society, 2006). The chances of getting breast cancer increases with age (Pike *et al.*, 1983; American Cancer Society, 2006a). The risks are also affected by a women's reproductive and menstrual history, especially the total accumulated exposure to ovarian oestrogen, early menarche and late menopause (Henderson at al., 1996; Kelsey and Bernstein, 1996; Colditz and Rosner, 2000). Women with a family history of the disease or who are carriers of the BRCA1 and BRAC2 genes have an increased risk (Henderson, *et al.*, 1996; Ponder, 2003). Other risk factors include: Race or ethnicity (National Cancer Institute, 2007a), history of prior breast biopsy (Fitzgibbons, 1998). Physical exercise (Friedenreich, 2001), exposure to radiation (National Cancer Institute, 2007a). The risk is increased in women who consume alcohol at harmful levels (Ridolfo and Stevenson, 2001). The use of hormonal replacement therapy (HRT) after menopause appears to increase the risk of getting breast cancer (Colditz *et al.*, 1995). However the issue of HRT as a risk factor is still under discussion (Malaysian Oncological Society, 2007).

1.3.6 Epidemiology of Breast Cancer

Breast cancer is the fifth leading form of cancer death worldwide where it accounted for 548,000 deaths (7% of all cancer deaths) worldwide in 2007. However, it is the most common form of cancer affecting women (American Cancer Society,

2006d; World Heath Organization, 2008). In most Western countries, it is the leading cause of cancer death in women between the ages 40 and 55 (Plant, 2001).

In the United States, breast cancer is the most common form of cancer in women, and the second most common cause of cancer death in women after lung cancer. In 2007, invasive breast cancer is expected to affect 178,480 women and cause 40,910 deaths or 7% of all cancer deaths in the U.S. (American Cancer Society, 2006d). Women in the U.S. and Europe have a 12.5% chance of developing invasive breast cancer in their lifetime and a 1 in 33 chance of breast cancer causing their deaths.

In Malaysia, 3738 female breast cancer cases were reported in 2003, making it the most diagnosed cancer among women. It accounted for 31% of newly diagnosed female cases. It is the most common form of cancer in all ethnic and age groups in women from the age of 15. Of the reported cases in 2003, 64.1% were in women between ages 40 and 60 years. The overall age specific rate (ASR) was 46.2 per 100,000 women. Of these, Chinese females had the highest ASR of 59.7 per 100,000 women, followed by Indian females with an ASR of 55.8 per 100,000 and Malay females with a 33.9 per 100,000 ASR. In 2003, the age pattern shown a peak age specific incidence rate of between ages 50-59 in Malays, Chinese and Indian, although this declined in older age groups (Lim and Halimah, 2004). About one in 19 women in Malaysia run the risk of developing breast cancer (Malaysian Oncological Society, 2007).

1.4 Proteomic Analysis and Protein Identification

Proteomics is the large-scale study of the proteome, particularly their structures and functions (Liebler, 2002). The concept of proteomics and proteome was first proposed by Wilkins in 1995 where the proteome is used to describe the entire complement of proteins expressed by a genome, cell, tissue or organism. Proteome analysis is a measurement of proteins in terms of their presence and relative abundance (Wilkins *et al.*, 1996). The ultimate aim of proteomics is to characterize protein pathways, networks and signaling events of cell regulation in diseases (Somiari *et al.*, 2004).

The study of proteomics can be divided into 3 main categories: Proteomic analysis, expression proteomics and cell-mapping proteomics. Proteomic analysis is the large-scale identification and characterization of proteins, including their post-translational modifications. Analyses are usually done using mass spectrometry. Expression proteomics uses proteomic tools such as two-dimensional gel electrophoresis and high performance liquid chromatography for global profiling of expressed proteins in cell lysates and tissues. As with proteomic analysis, analyses of protein identities are done by mass spectrometry. Cell-mapping proteomics utilizes the identification of protein complexes to determine protein to protein interactions and intracellular signaling circuitry (Simpson, 2003).

In current practice, proteomics encompasses four main applications: Proteome mining, protein expression profiling, identification of protein-protein interactions in protein complexes, and protein modification profiling. Proteome mining attempts to identify as many components of the proteome as possible. Protein expression profiling aims to detect and measure differences in protein expression between samples. It is used to identify proteins that are expressed

normally and in the presence of a disease in a living system. In protein-protein interactions identification, the identity and function of the individual proteins are determined before working out how they interact with one another. The purpose of profiling protein modifications is to establish when proteins are modified after their synthesis (Liebler, 2002; Simpson, 2003).

Mass spectrometry (MS) based methods (Mann and Talbo, 1996) for the identification of gel-separated proteins have become an essential tool in proteomics. It has replaced Edman's degradation as a protein identification tool as it is more sensitive, has higher throughput and able to identify proteins in the presence of other proteins (Simpson, 2003). Current MS based methods rely on the separation of proteins with either one-dimensional (Laemmli, 1970) or two-dimensional electrophoresis (O'Farrell, 1975) followed by proteolytic digestion of the proteins (Schevchenko *et al.*, 1996). The peptide mass or the sequence information from fragmented peptides is usually sufficient to allow unambiguous protein identification (Simpson, 2003).

1.4.1 Sample Preparation

Sample preparation is one of the most critical steps in proteomics (Link, 1999; Rabilloud, 1999; Molloy, 2000) and in producing optimal and reproducible results in proteomics (Dunbar, 1987). It is perhaps the most important step in proteome analysis as this step will influence protein yield (Simpson, 2003; Dhingra *et al.*, 2005). An effective sample preparation will yield a high concentration of proteins as well as free of salts and other interfering substances such as nucleic acids and lipids which can interfere with subsequent protein analysis (Jiang *et al.*, 2004).

Ideally, the process will result in the complete solubilization of proteins in the extraction buffer (Rabilloud, 1996).

Sample preparation usually starts with tissue lysis or disruption, which should be done in a solubililizing solution. The sample can be precipitated to remove interfering substances before being resolubilized in the solubilizing solution. In general, it is best to keep sample preparation as simple as possible (Simpson, 2003).

1.4.2 Tissue Disruption and Sequential Extraction

The first step of a sample preparation protocol is to wash the tissue, disrupt the tissue in a suitable buffer using homogenizer. The homogenate is centrifuged to separate the pellet from the supernatant. There are many methods to homogenize tissues but the ultimate purpose is to break the tissue or cell in order to release proteins (Görg *et al.*, 2000). The centrifugation process will separate soluble proteins in the supernatant from the insoluble proteins and cell debris that remain as pellet. A stronger solubilizing solution is then added to solubilize the remaining proteins in the pellet. This method is known as sequential extraction where proteins are sequentially extracted in increasingly powerful solubilizing solutions (Molloy *et al.*, 1998; Simpson, 2003).

1.4.3 Precipitation

Samples usually contain contaminating substances such as salts, detergents, nucleic acids and lipids. These contaminants can interfere with subsequent proteomic analysis such as two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). Samples are precipitated and centrifuged. The contaminants remain soluble in the supernatant will be removed by discarding the supernatant. The

protein pellet is resuspended in a solubilizing solution. This has an added advantage of concentrating the protein (Simpson, 2003) and inactivating proteases (Rabilloud and Chevallet, 2000). One of the methods used to precipitate proteins is by trichloroacetic acid/acetone (TCA/A) precipitation that is known for its high protein recovery and ease of use (Görg *et al.*, 1988a, 1997; Tsugita and Kamo, 1999; Jiang *et al.*, 2004).

1.4.4 Sobulization

Solubilizing solutions are a mixture of chemicals to fully solubilize proteins. These chemicals must not interfere with 2D-PAGE and also keep the proteins in their solubilized state during 2D-PAGE runs. In sequential extractions, the first solubilizing solution is Tris while the subsequent solution is thiourea lysis buffer (TLB), which has a comparatively higher solubilizing properties. TLB contains chemicals such as chaotropic agents, detergents, reducing agents and ampholytes (Molloy *et al.*, 1998).

Tris [tris(hydroxymethyl)aminomethane] is widely used in biochemistry and molecular biology as a buffer (Gomori, 1955). A buffer is a substance where its presence resists changes in pH by increasing the amount of acid or alkali required for the change (van Slyke, 1922; Campbell and Farrell, 2006). Tris is slightly basic with a pKa of 8.08 (Bates and Pinching, 1949) which makes it suitable for studies in biological systems as well as pH control *in vitro* (Nahas, 1961; Bates *et al.*, 1978). It is used at a concentration of 40mM to solubilize aqueous soluble proteins (Molloy *et al.*, 1998; Santoni *et al.*, 2000).

Chaotropic agents denature proteins by disrupting their hydrogen bonds. The disruption prevents unwanted protein aggregation or formation of secondary

structures that affect protein mobility. Urea is one of the most widely used chaotropic agents. It maintains proteins in their denatured state, keeps them soluble and unfolds most proteins to their fully random conformation (Dunbar, 1987). Urea is typically used at a concentration of 8M. However, as some proteins still do not get denatured at high urea concentrations (Tanford, 1968), the addition of thiourea can improve the solubility of proteins, particularly membrane proteins (Pasquali *et al.*, 1997; Rabilloud *et al.*, 1997; Rabilloud, 1998).

Detergents ensure complete sample solubilization and prevent protein aggregation by disrupting protein hydrophobic interactions. They also increase protein solubility at their respective isoelectric point (pI). For proteins to be separated by 2D-PAGE, the detergents used must be non-ionic or zwitterionic to allow proteins to migrate according to their own charges (Berkelman and Stenstedt, 1998). The use of detergents is quite important as some proteins, especially membrane proteins require the presence of detergents to sustain their solubility during isolectric focusing (IEF) (Garfin and Heerdt, 2000).

Earlier use of non-ionic detergents in 2D-PAGE (Klose, 1975; O'Farrell, 1975) has been superseded by the use of zwitterionic cholamidosulfobetaine detergents 3-[3-cholamidopropyl]dimethylammonio]-1-propane sulfonate (CHAPS) and its hydroxyl analog 3-[3-cholamidopropyl]dimethylammonio]-2-hydroxy-1-propane sulfonate (CHAPSO) (Perdew *et al.*, 1983). CHAPS is a non-denaturing detergent for solubilizing proteins and is available in high purity and can be used at concentrations up to 4% in isoelectric focusing (IEF) (Garfin, 1990; Herbert, 1999; Simpson, 2003). It is especially suitable for protein solubilization in 2D-GE (Hjelmel and Chrambach, 1981; Garfin, 1990; Schupbach *et al.*, 1991). Other

detergents such as SB-10 are also used (Adessi *et al.*, 1997; Molloy *et al.*, 1998) but they are relatively insoluble in the presence of urea. A typical concentration of detergents in solubilization buffers is commonly in the range of 0.5-4%.

During IEF, some proteins tend to precipitate at their pI even in the presence of chaotropic agents and detergents. Therefore, carrier ampholytes are added to improve protein solubility by maintaining the optimum pH for maximum protein solubility and by minimizing protein aggregation due to charge-charge interactions (Garfin and Heerdt, 2000; Simpson, 2003). Carrier ampholytes also ensure uniform conductivity during IEF without altering the pH gradient of the IPG strip. The concentration of carrier ampholytes in sample preparation is typically 0.2-0.4% (w/v).

Reducing agents are used in solubilization buffers to cleave disulfide bonds between cysteine residues and to maintain proteins at their reduced state (Liu, 1977). This allows proteins to be analyzed as single subunits (Garfin and Heerdt, 2000). Originally, 2-mercaptoethanol was used to reduce samples for 2D-GE, but because it has some disadvantages (Marshall and Williams, 1984), it has been superseded by the sulfhydril reducing agent 1,4 – dithiothreitol (DTT) (Cleland, 1964). The concentration of DTT used is between 20 to 100mM but 50mM is usually effective in breaking disulfide bonds (Garfin and Heerdt, 2000).

Homogenization of tissue samples not only releases proteins but also proteases. These proteases can degrade proteins and complicate analysis of 2D-PAGE results (Berkelman and Stenstedt, 1998; Simpson, 2003). Protease inhibitors inhibit the activity of proteases. Phenyl methyl sulphonyl fluoride (PMSF) and AEBSF [4-(2-aminoethyl) benzylsulfonyl fluoride] are examples of inhibitors of serine and cysteine proteases (Mintz, 1993; Simpson, 2003). They inhibit the

proteases' function by forming covalent bonds with these proteases. Typical concentrations used are between 1 to 10mM.

1.4.5 Electrophoresis

Electrophoresis refers to the transport of charged particles or molecules through a solvent or matrix in an electric field towards their opposing charges (Dunbar, 1987; Campbell and Farrell, 2006). The separation of molecules not only depends on the size, shape and net charge of the molecule, but also by voltage, electrode distance, medium viscosity, temperature and time (Simpson, 2003). Since different proteins have different shapes and net charges, each protein will migrate in an electrophoretic field at a different rate (Dunbar, 1987). In general, proteins with a positive and negative net charge migrate to the cathode and anode electrodes respectively, while neutral charged proteins do not move at all (McKee and McKee, 1999).

The use of electrophoresis to resolve proteins is widely used in biochemistry and proteomics due to its resolving power, versatility and adaptability (Figeys *et al.*, 1998). Not only does it separate proteins, it allows crucial protein properties such as isoelectric point (pI) and molecular weight (MW) to be determined (Nelson and Cox, 2006). Most electrophoretic procedures today are based on zone electrophoresis (Hames, 1990) or discontinuous electrophoresis (Davis, 1964; Ornstein, 1964) in polyacrylamide gels (Raymond and Weintraub, 1959).

1.4.6 Polyacrylamide Gel Electrophoresis

Polyacrylamide is one of the most widely used matrices for carrying out protein seperation by electrophoresis (Dunbar, 1987). The small pore sizes of its matrix give polyacrylamide gels molecular sieving properties. This allows polyacrylamide to separate proteins based on their shape and size in addition to charge density (Hames, 1990). Polyacrylamide gel electrophoresis (PAGE) is one of the main methods used to separate proteins in complex mixtures and determining protein purity. It can also be used to find out a protein's relative molecular mass, verify its concentration and detect protein modifications (Simpson, 2003).

A polyacrylamide gel is formed by the free-radical vinyl polymerization of the monomer acrylamide into long chain polymers and cross-linking these by bifunctional compounds such as N,N'-methylene bisacrylamide (bisacrylamide) that react with free functional groups at the ends of the chain (Richards and Lecanidou, 1974). The polymerization of acrylamide is initiated by a chemical peroxide, usually ammonium persulfate (APS). APS dissolved in water forms persulfate radicals that acts as an electron carrier that provides unpaired electrons and converts the acrylamide monomer to a free radical state. The free radical acrylamide monomer then reacts with other acrylamide monomers to begin the chain elongation and acrylamide polymerization (Shi and Jackowski, 1998). Tertiary or quaternary amines such as N,N,N',N'-tetramethylenediamine (TEMED) are added to accelerate the formation of free radicals, which in turn accelerates the polymerization process The elongating chains are crosslinked with bisacrylamide to form a matrix with pore sizes depending on the acrylamide monomer concentration (Hames, 1990; Dunbar, 1987; Simpson, 2003). The crosslink adds rigidity and tensile strength to the gel

(Sambrook and Russell, 2001). Bisacrylamide is an important component as without it the acrylamide polymer chains will not gel (Cooper, 1977).

The composition of a polyacrylamide gel is described by the monomer concentration, %T and crosslinker concentration, %C (Hjertén, 1962). Generally, the pore size of polyacrylamide gels is inversely proportional to %T (Hames, 1990).

Polyacrylamide gels are the separation media in isoelectric focusing (IEF) and sodium dodecyl sulfate polyacrlyamide gel electrophoresis (SDS-PAGE). IEF and SDS-PAGE are components of two-dimensional gel electrophoresis (2D-GE), where proteins are separated by two different parameters, i.e. pI and MW. 2D-PAGE is one of the most effective and powerful means of resolving and analyzing complex mixtures. (Celis and Bravo, 1984; Dunbar, 1987; Liebler, 2002; Simpson, 2003).

1.4.7 Isoelectric Focusing

Proteins are amphoteric molecules that carry either a positive, negative or zero net charge depending on the pH of their surroundings (Simpson, 2003). The net charge of a protein molecule depends on the sum of all charges of its amino acid side chains (Dunbar, 1987). At a specific pH, the net charge of a protein will be zero, and this pH value is referred to as the protein's pI.

Isoelectric focusing (IEF) is a form of electrophoresis where proteins are separated based on their pI. A protein is placed in a medium with a pH gradient and is subjected to an electric field at the same time. In the presence of a pH gradient, a protein will migrate to the position in the gradient equal to its pI. A protein that migrates nearer to its pI becomes less charged until it stops at its pI as it has no net charge. This is the focusing effect of IEF, which concentrates proteins at their pI and allows separation of proteins on the basis of very small differences in net charge