DRUG DISCOVERY TARGETING PROTEOMIC APPROACH

A Dissertation submitted to the Department of Pharmacy, East West University, in partial fulfillment of the requirements for the degree of Masters of Pharmacy.

Submitted By

Samiya Khondaker Rinta

ID: 2014-3-79-024

Research Invigilator

Dr. Repon Kumer Saha

Assistant Professor

Department Of Pharmacy

East West University

Dedication

This Research Paper is dedicated to

My beloved parents,

Who are my biggest Inspirations…

DECLARATION BY THE CANDIDATE

I, Samiya Khondaker Rinta, hereby declare that this dissertation, entitled **"Drug discovery targeting proteomic approach"** submitted to the Department of Pharmacy, East West University, in the partial fulfillment of the requirement for the degree of Masters of Pharmacy, is a genuine & authentic research work carried out by me under the guidance of Dr. Repon Kumer Saha, Assistant Professor, Department of Pharmacy, East West University, Dhaka. The contents of this dissertation, in full or in parts, have not been submitted to any other Institute or University for the award of any Degree or Diploma of Fellowship.

Samiya Khondaker Rinta

ID: 2014-3-79-024

Department of Pharmacy

East West University

Aftabnagar, Dhaka

CERTIFICATION BY THE SUPERVISOR

This is to certify that the dissertion, entitled **"Drug discovery targeting proteomic approach"** is a bonafide research work done, under our guidance and supervision by Samiya Khondaker Rinta (ID: 2014-3-79-024), in partial fulfillment of the requirement for the degree of Masters of Pharmacy.

Dr. Repon Kumer Saha

Assistant Professor

Department of Pharmacy

East West University

Aftabnagar, Dhaka

ENDORSEMENT BY THE CHAIRPERSON

This is to certify that the dissertation, entitled **"Drug discovery targeting proteomic approach"** is a bonafide research work done by Samiya Khondaker Rinta (ID: 2014-3-79-024), in partial fulfillment of the requirements for the degree of Masters of Pharmacy.

Dr. Shamsun Nahar Khan

Chairperson & Associate Professor

Department of Pharmacy

East West University

Aftabnagar, Dhaka

ACKNOWLEDGEMENTS

All praise is for Almighty for all the bounties granted to me and only with His guidance and help this achievement has become possible.

It is my pleasure and proud privilege to express my heartiest regards and gratitude to my respected teacher and supervisor **Dr. Repon Kumer Saha**, Assistant Professor, Department of Pharmacy, East West University, for his expert supervision, constructive criticism, valuable advice, optimistic counseling, constant support and continuous backup and encouragement throughout every phase of the project as well as to prepare this dissertation.

I would also like to put forward my most sincere regards and profound gratitude to **Dr. Shamsun Nahar Khan**, Chairperson & Associate Professor, Department of Pharmacy, East West University, for giving me the opportunity to conduct such an interesting project and for facilitating a smooth conduction of my study

I would also like to extend my thanks to all the research students in the lab, lab officers and other staffs of the Department of Pharmacy for their help and assistance, friendly behavior and earnest co-operation which enabled me to work in a very congenial and comfortable ambience.

I owe special thanks to my fellow research group members for their immense support and contribution in my research work.

Last but not the least, I would like to thank my family, and friends for their care and encouragement during my research work.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

2DE	Two-Dimensional Electrophoresis
2DGE	Two-dimensional gel electrophoresis
2D-DIGE	Two-dimensional fluorescence difference gel electrophoresis
AMPKα1	AMP-activated protein kinase α1
APL	Acute Promyelocytic Leukemia
bFGF	Basic fibroblast growth factor
BTK	Bruton's tyrosine kinase
CAMK2G	Calcium/calmodulin-dependent protein kinase type II gamma chain
CDKs	Cyclin-dependent kinases
CK1	Casein kinase 1
CML	Chronic Myologenous Leukemia
CMML	Chronic Myolomonocytic Leukemia
CRBN	Cereblon
DDR1	Discoidin domain receptor-1
DIGE	Difference Gel Electrophoresis
ESI	Electrospray ionization

FGFR	Fibroblast growth factor receptor
FTIC	Fourier transform ion cyclotron
GAK	Cyclin G-associated kinase
GI	Gastrointestinal
GIST	Gastrointestinal Stromal Tumour
HPLC	High performance liquid chromatography
ICAT	Isotope-Coded Affinity Tag
IEF	Isoelectric focusing
iTRAQ	Isobaric Tag for Relative and Absolute Quantitation
JAK1	Janus kinase 1
LC	Liquid chromatograph
MALDI	Matrix assisted laser desorption/ionization
MS	Mass spectrometry
MUDPIT	Multidimensional Protein Identification Technology
NQO2	Quinone oxidoreductase 2
PDGFR	Platelet-Derived Growth Factor receptor
PDXK	Pyridoxal kinase

PKC	Protein kinase C
PKNβ	Protein kinase N beta
PTMs	Post-translational modifications
RICK	Rip-like interacting caspase-like apoptosis-regulatory protein kinase
RP	Reversed phase
SCX	Strong cation exchange
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SELDI	Surface-enhanced laser desorption/ionization
SILAC	Stable Isotope Labeling with Amino Acids in Cell Culture
TOF	Time-of-flight
VDAC	Voltage-dependent anion channel
ULK3	Unc-51-like kinase 3

ABSTRACT

The drug discovery process is not a limited or established series of steps. Modern approaches include target-based drug discovery in which researchers require to survey proteins. The two most important requirements for this type of technology are to find more effective biomarkers for detecting disease and to discover proteins to which therapeutic drugs can be targeted. It is well-known that the risks in drug discovery process are high and it requires a great deal of time before it is known whether a target candidate drug will succeed or fail. The key to success for pharmaceutical companies is to make accurate decisions using an accelerated process. Genomics revolution had a very positive impact upon these issues.

Proteomics is in the field, now, as a powerful new partner of genomics. It will provide a significant new aspect to drug discovery by providing the potential to analyze proteins from a very wide diversity of biological systems in a high-throughput and systematic way. Each step of the process from target discovery to clinical trials is accessible to proteomics. Scientists are able to see every detail of their biological focus, from genes, mRNA, proteins and their subcellular localization. This will greatly help our knowledge of the fundamental mechanistic basis of human disease and will assist in the discovery of improved, faster, minimum toxic and hopefully, inexpensive drugs.

Key words: Genomics, proteomics, genes, mRNA, biomarkers, proteins.

Chapter One
INTRODUCTION

1.1 Proteins

Proteins are large and complex biological molecules. They are composed of carbon, hydrogen, nitrogen, and oxygen, which are arranged as hundreds or thousands of monomer units called amino acids. They are involved in many critical roles in the body. They required for the regulation, structure, and function of the body's organs and tissues (Branden and Tooze, 1998; Nelson and Cox, 2008).

Table 1: Example of some proteins and their function (Branden and Tooze, 1998; Nelson and Cox, 2008).

Example	Function	Description
Growth hormone	Messenger	Transmit signals to coordinate biological processes between different cells, tissues, and organs.
Phenylalanine hydroxylase	Enzyme	Carries out many chemical reactions in cells and assists formation of new molecules.
G Protein-coupled receptor	Receptor	Has a ligand-binding site on the cell surface and an effector domain within the cell, which may undergo a conformational change or may have an enzymatic activity.
IgG (Immunoglobulin G)	Antibody	Binds to specific foreign particles (viruses and bacteria) and helps to protect the body.
Potassium channel	Channel	Allows molecules to pass through the cell membrane
Haemoglobin	Transport	Binds with oxygen molecules and transport them to other locations in the organism.
Actin	Structural component	Provides structure and support for cells.

1.2 Amino acids

Amino acids are fundamental building blocks of protein. A protein is formed through combination of twenty different types of amino acids. The unique three-dimensional structure and the specific function of a protein are determined by the amino acid present and their

sequence. Amino acids consist of carbon, nitrogen, hydrogen, oxygen and, sometimes, sulfur and phosphorous. Large protein complexes are called polypeptides, which are made of long chains of amino acids (Berg *et al.*, 2002; Wu, 2013).

The general structure of amino acids consists of a carbon center and its four substituents, which consists of an amino group (NH_2), an organic acid (carboxyl) group (COOH), a hydrogen atom (H), and a fourth group, referred to as the R-group, that determines the structural identity and chemical properties of the amino acid. The first three groups are common to all amino acids. The basic amino acid structure is R-CH(NH_2)-COOH (Berg *et al.*, 2002; Wu, 2013).

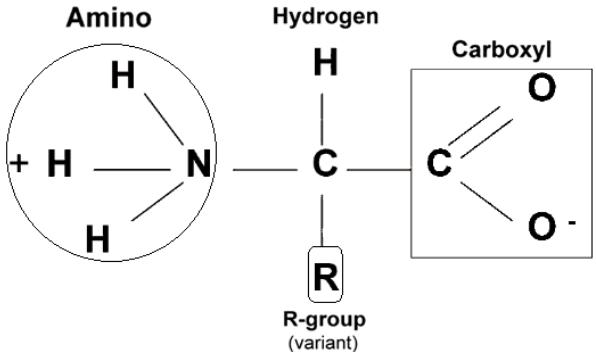
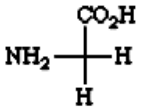
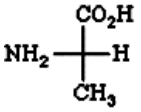
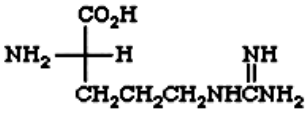
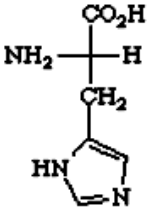
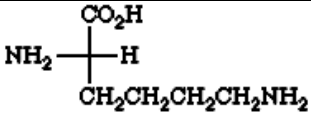
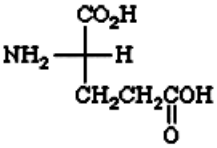
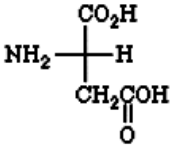
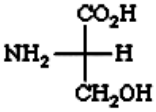
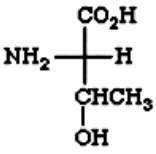


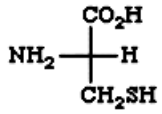
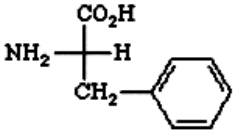
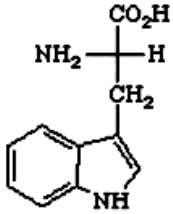
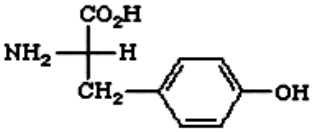
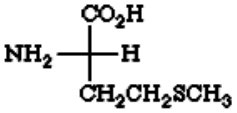
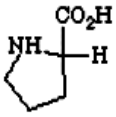
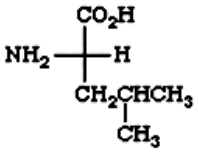
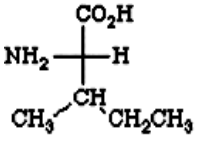
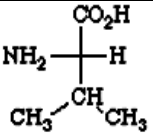
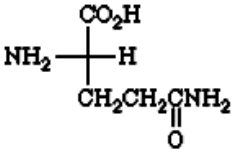
Figure 1: General structure of amino acid (Berg *et al.*, 2002; Nelson and Cox, 2008; Wu, 2013).

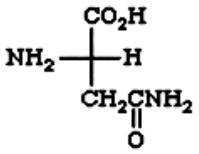
Its unique side chain (R-group) distinguishes one amino acid from another and dictates an amino acid's chemical properties. Since the amino acids contain an amino group (base) at one end and a carboxyl group (acid) at another end, a proton can be transferred from -COOH to -NH_2. The compound containing an ammonium cation and a carboxylate anion is called a zwitterion (Berg *et al.*, 2002; Wu, 2013).

Humans do not have all the enzymes required for the biosynthesis of all the twenty amino acids. The 10 of the 20 amino acids that humans can produce are alanine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, proline, serine and tyrosine. Tyrosine is produced from phenylalanine, so tyrosine is required if the diet is deficient in phenylalanine. The essential amino acids are arginine (required for children and not for adults), histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine. These amino acids are required in the diet. Plants are able to make all the amino acids (Berg *et al.*, 2002; Wu, 2013).

Table 2: Twenty types of amino acids (Alberts *et al.*, 2002; Colorado State University, 2002; Nelson and Cox, 2008; Atlas of Genetics and Cytogenetics in Oncology and Haematology, 2015).

Type	Name	Abbreviation	Structure
Small	Glycine	Gly, G	
	Alanine	Ala, A	
Basic	Arginine	Arg, R	
	Histidine	His, H	
	Lysine	Lys, K	
Acidic	Glutamic acid	Glu, E	
	Aspartic acid	Asp, D	
Nucleophilic	Serine	Ser, S	
	Threonine	Thr, T	

Type	Name	Abbreviation	Structure
	Cysteine	Cys, C	
Aromatic	Phenylalanine	Phe, F	
	Tryptophan	Trp, Y	
	Tyrosine	Tyr, W	
Hydrophobic	Methionine	Met, M	
	Proline	Pro, P	
	Leucine	Leu, L	
	Isoleucine	Ile, I	
	Valine	Val, V	
Amide	Glutamine	Gln, Q	

Type	Name	Abbreviation	Structure
	Asparagine	Asn, N	

1.3 Peptides

Amino acids combine to form amides called peptides. The -CO-NH- bond or linkage is called the peptide bond. The amino acids when linked by peptide bonds are referred to as residues. If glycine on the left and alanine on the right come together, they can form a dipeptide. The nitrogen on the amino group of the alanine forms a bond with the carbonyl carbon of the carboxylic acid on the glycine. The byproduct of this reaction is water. So this is a dehydration reaction, or loss of water reaction. The resulting molecule that's formed is a dipeptide, or peptide consisting of two amino acids (Berg *et al.*, 2002; Nelson and Cox, 2008; Wu, 2013).

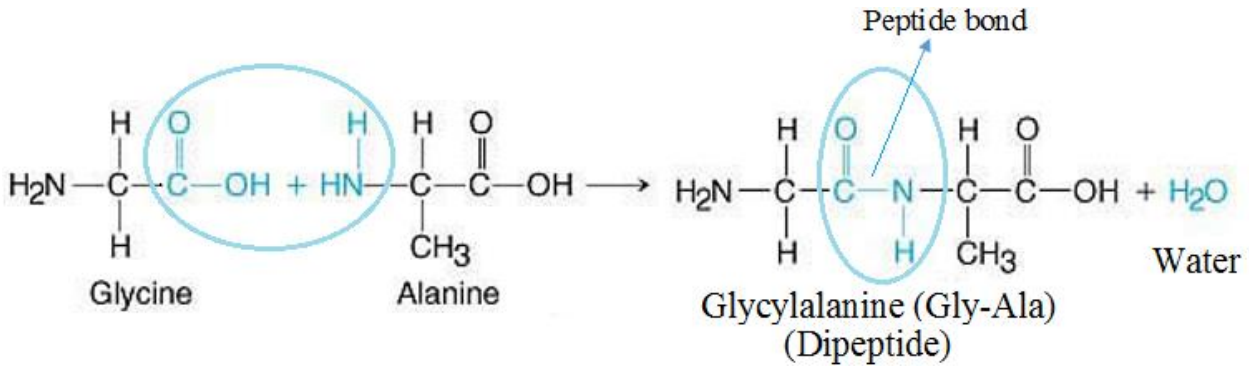


Figure 2: Dipeptide formation (Berg *et al.*, 2002; Nelson and Cox, 2008; Wu, 2013).

If this dipeptide comes into contact with serine, it is possible that it could form a tripeptide when the nitrogen of the serine amino acid bonded to the carbonyl carbon of the carboxylic acid on the alanine in the dipeptide. The byproduct of this dehydration reaction is water, and the resulting product is a tripeptide, made up of glycine, alanine, and serine. The amino acid sequence of a protein is always presented in the N-to-C direction, reading from left to right. The end of the peptide chain with the free amino group is known as the N-terminus (or amino-terminus). The end of the peptide chain with the free carboxylic acid group is known as the C-terminus (or carboxyl-terminus) (Berg *et al.*, 2002; Nelson and Cox, 2008; Wu, 2013).

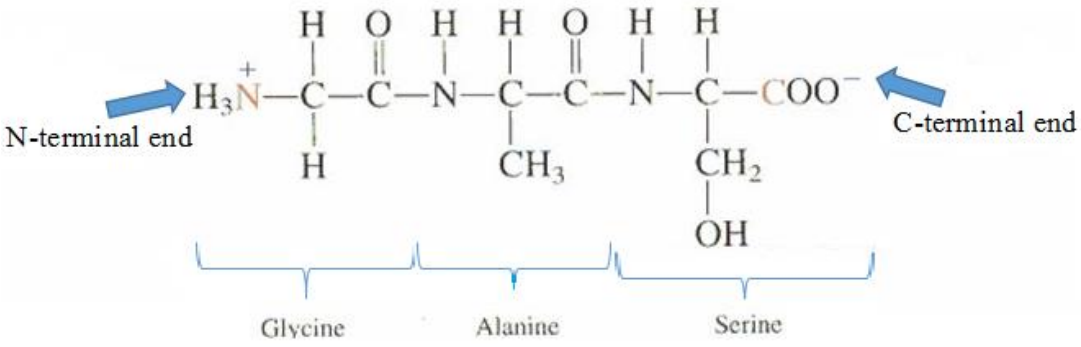


Figure 3: Tripeptide formation (Berg *et al.*, 2002; Nelson and Cox, 2008; Wu, 2013)

Many amino acids form a polypeptide. Polypeptides involving more than about hundred amino acids are called proteins. When hundred or more amino acids are involved in the production of a protein and each of these amino acids may have one of twenty different structures, the number of possible proteins is huge (Berg *et al.*, 2002; Nelson and Cox, 2008; Wu, 2013).

1.4 Levels of protein structure

The structural features of proteins are normally described at four different levels of complexity:

Primary structure: This describes the unique linear arrangement by which the amino acids are linked together to form a protein or polypeptide chain and the location of covalent linkages such as disulfide bonds between amino acids. The amino acid sequence of a protein is determined by the information found in the cellular genetic code. As discussed before that a particular protein has a unique and specific arrangement of amino acids in a polypeptide chain, so altering a single amino acid can cause a gene mutation, often resulting in a non-functioning protein (Branden and Tooze, 1998; Berg *et al.*, 2002; Nelson and Cox, 2008).

Figure 4: Primary structure of protein (Berg *et al.*, 2002; Nelson and Cox, 2008; Wu, 2013)

Secondary structure: After synthesis of the primary structure, polypeptide chains are folded or pleated into different shapes, called their secondary structure. Two common examples of secondary structures include alpha helices and beta pleated sheets, which are stabilized by hydrogen bonding and this gives the shape great stability (Branden and Tooze, 1998; Nelson and Cox, 2008).

- **Alpha (α) helix structure:** This structure looks like a coiled spring and is secured by hydrogen bonding in the polypeptide chain.

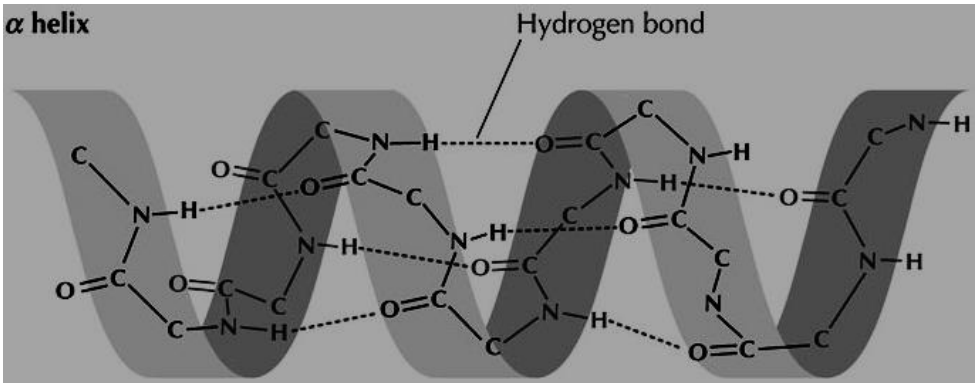


Figure 5: Structure of alpha helix (Branden and Tooze, 1998; Berg *et al.*, 2002; Nelson and Cox, 2008).

Beta (β) pleated sheet: This structure is folded or pleated and is held together by hydrogen bonding between polypeptide units of the folded chain that lie adjacent to one another (Branden and Tooze, 1998; Nelson and Cox, 2008; Wu, 2013).

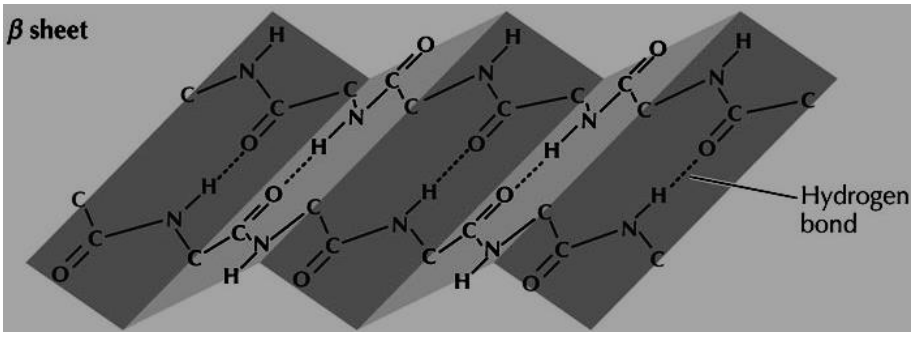


Figure 6: Structure of beta pleated sheet (Branden and Tooze, 1998; Berg *et al.*, 2002; Nelson and Cox, 2008).

Tertiary structure: The final three-dimensional structure of a protein is its tertiary structure, which results from a large number of non-covalent interactions between amino acids. This

involves coiling or pleating. There are four different bonds and interactions that hold a protein in its tertiary structure (Branden and Tooze, 1998; Berg *et al.*, 2002; Nelson and Cox, 2008).

- **Disulphide Bonds** - Where two cysteine amino acids are found together, a strong double bond (S=S) is formed between the sulphur atoms within the Cysteine monomers.
- **Ionic Bonds** - If two oppositely charged 'R' groups (+ve and -ve) are found close to each other, and ionic bond forms between them.
- **Hydrogen Bonds** – It is the electrostatic interaction between the amide and carbonyl groups of the amino acids.

Hydrophobic and Hydrophilic Interactions - Some amino acids may be hydrophobic while others are hydrophilic. In a water based environment, a globular protein will orientate itself such that its hydrophobic parts are towards its center and its hydrophilic parts are towards its edges (Branden and Tooze, 1998; Berg *et al.*, 2002; Nelson and Cox, 2008).

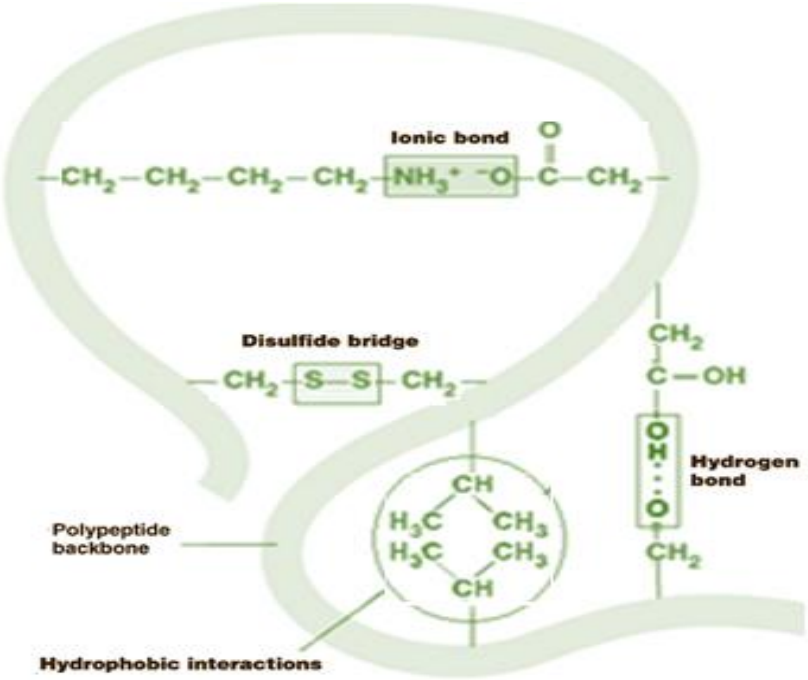


Figure 7: Secondary structure of protein (Branden and Tooze, 1998; Berg *et al.*, 2002; Nelson and Cox, 2008; Wu, 2013)

Tertiary structure can be broken by the action of heat. Increasing the kinetic energy causes a tertiary structure to vibrate more, as a result the bonds (which are mainly weak, non-covalent bonds) will be more likely to break and will not be able to that maintain its shape. When a

protein loses its shape in this way it is said to be denatured. Even cooling the protein will not return it to its original complex form (Branden and Tooze, 1998; Nelson and Cox, 2008; Wu, 2013).

Proteins with a three dimensional structure fall into two major types:

- **Globular** - These proteins tend to form ball-like structures with their hydrophobic parts towards the center and their hydrophilic towards the edges. This makes them water soluble. They mostly have metabolic functions, for example, plasma proteins, antibodies in mammals and enzymes in all organisms.
- **Fibrous** - These proteins form long fibers and are mostly consist of repeated sequences of amino acids which are insoluble in water. They usually have structural function, for example, keratin found in nails and hair, collagen found in cartilage and bone (Branden and Tooze, 1998; Berg *et al.*, 2002; Nelson and Cox, 2008).

Quaternary structure: non-covalent interactions that bind multiple polypeptides into a single, larger protein. Hemoglobin has quaternary structure due to association of two alpha globin and two beta globin polyproteins. This refers to the structure of a protein macromolecule formed by interactions between multiple polypeptide chains. Each polypeptide chain is referred to as a subunit. Proteins with quaternary structure may consist of more than one of the same type of protein subunit. They may also be composed of different subunits. An example of a protein with quaternary structure is hemoglobin. It is found in the blood. It is an iron containing protein that binds oxygen molecules. It has four subunits which are two alpha subunits and two beta subunits (Branden and Tooze, 1998; Berg *et al.*, 2002; Nelson and Cox, 2008; Wu, 2013).

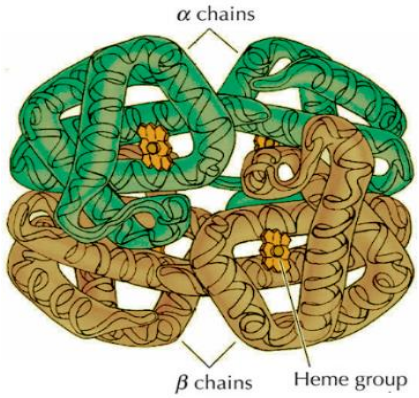


Figure 8: Hemoglobin with four sub-units (Berg *et al.*, 2002; Nelson and Cox, 2008).

1.5 Protein synthesis

New proteins are continuously produced to replenish older molecules that have been degraded, throughout the lifetime of a cell. The genetic information required for protein synthesis is stored in the DNA Genes (units of DNA) encodes information needed to produce a specific protein molecule and the genes are assembled together in a strand of DNA and associated proteins called chromosomes (Alberts *et al.*, 2002; Nelson and Cox, 2008).

The first step in new protein synthesis is transcription, which is synthesis of a messenger RNA (mRNA) that exactly complements the sequence of the DNA of a particular gene. Here one strand of the DNA double helix is used as a template. Once the mRNA molecule is synthesized, it is then sent to the cytoplasm, where all the components (amino acids, ribosomes, transport RNAs, etc.) required for the actual protein synthesis are brought together (Alberts *et al.*, 2002; Nelson and Cox, 2008).

Each mRNA encodes the information for a single protein. The ribosome consists of proteins and ribosome RNA molecules (rRNA), which appears in two subunits. The mRNA initially binds to one of the ribosomal sub-units. When the mRNA interacts with the big ribosome sub-unit, this triggers the approach of another RNA molecule, called transfer RNA (tRNA). The tRNA molecule possess a specific sequence of 3-bases (anti-codon), which must complement to a corresponding sequence (codon) within the mRNA sequence. When it finds it, it attaches to the mRNA, as the other end of the tRNA is "loaded" with an amino acid. The first tRNA binds to a so called "start codon", which is one and the same for all proteins. As the complete ribosome structure is formed, another tRNA molecule approaches. The next tRNA differ from the first one and is carrying another amino acid. Again, the tRNA must have an anti-codon that matches complementary the second codon of the mRNA. The two amino acids carried by the first two tRNAs are bind together with help from the ribosome and using cellular energy in the form of adenosine triphosphate (ATP) (Alberts *et al.*, 2002; Nelson and Cox, 2008).

The above steps repeats until there are uncoupled codon sequences on the mRNA - thus the chain of amino acids grows longer. Once the sequence of amino acids is successfully assembled in a protein, the two ribosome sub-units separate from each other, to be joined again for later use.The actual sequence of amino acids forms the so called primary structure of the proteins. Depending

on the exact composition and order of the amino acids in the protein sequence, the chain folds into a three-dimensional shape. When this happens the protein is complete (Alberts *et al.*, 2002; Nelson and Cox, 2008).

The process of protein synthesis takes place in multiple ribosomes simultaneous and all throughout the cell cytoplasm. A living cell can synthesize hundreds of different proteins every single second. The mRNA molecule, in short, is used as a template by the ribosome of the cell to produce a protein molecule having the exact amino acid sequence that is present in the ribo-nucleotide sequence of mRNA molecule that were complementary to the sequence of deoxyribo-nucleotides of the DNA forming the gene. This process of formation of a protein molecule from the mRNA molecule is called translation. Thus, the overall process of new protein synthesis, or gene expression, involves two key steps:

1. Transcription of the DNA sequence to a messenger RNA (mRNA), and
2. Translation of the mRNA structure into the amino acid sequence of a protein product (Branden and Tooze, 1998; Alberts *et al.*, 2002; Nelson and Cox, 2008).

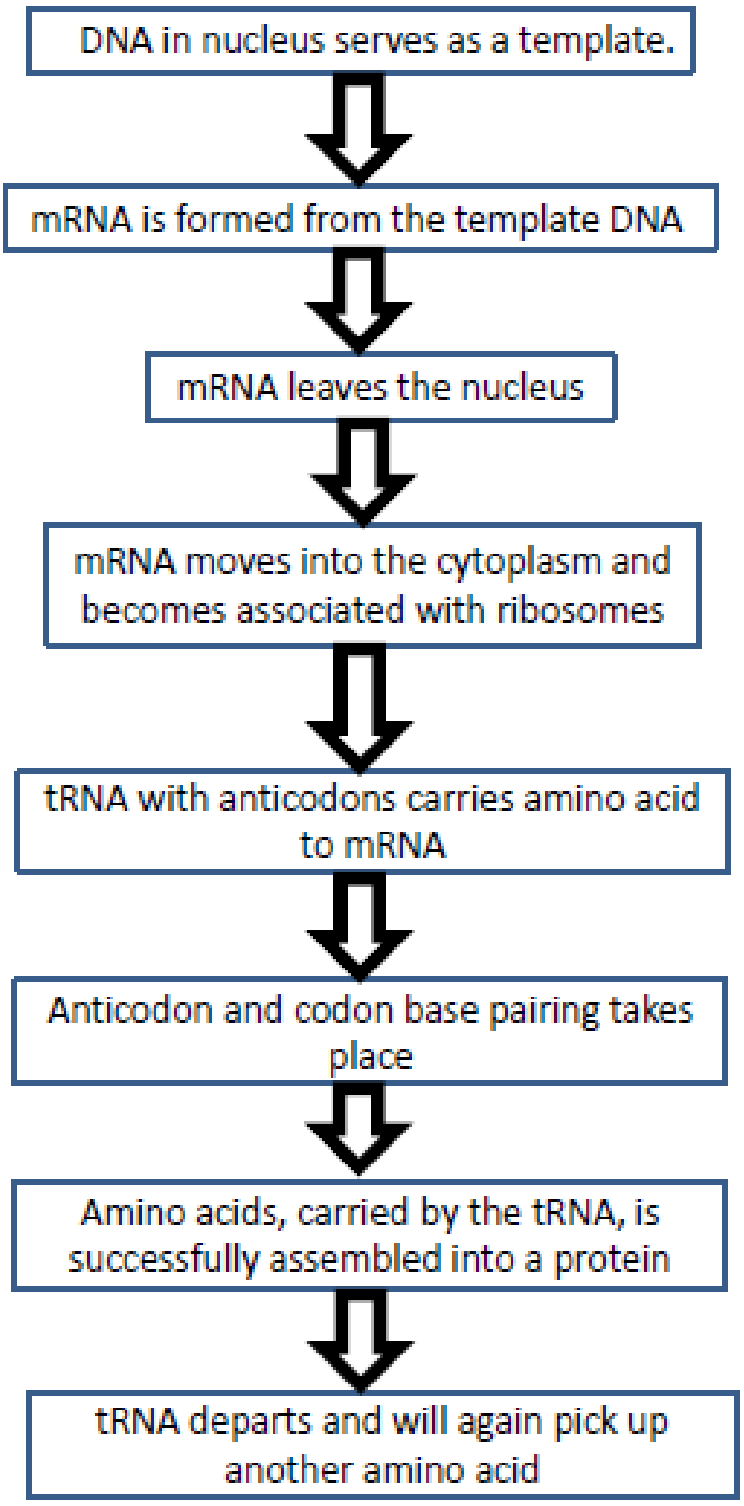


Figure 9: Steps of protein synthesis (Branden and Tooze, 1998; Alberts *et al.*, 2002; Nelson and Cox, 2008).

1.6 Drug discovery using proteomic approach

Drug discovery is an inherently complex process with a history spanning thousands of years. However, the elusive mechanisms of action and limited specificity of compounds hamper their further application in clinical practice. A large number of drugs have been found to act upon multiple targets, inevitably resulting in side effects and drug resistance during treatment. The most notorious drug in history, thalidomide, which was used to alleviate morning sickness during pregnancy, was found to cause fetal malformations and multiple birth defects. In addition, the difficulty in matching numerous complicated drugs with the desired physiologic effects is best characterized by the endless battle against drug resistance in antibiotics as well as in anticancer therapies (Huang, *et al.*, 2012).

It has been appreciated that the more we grasp the molecular mechanisms of potent drugs, the more we realize that these drugs are unexpectedly promiscuous to their relevant targets and the more we become aware of the obstacles lying ahead. However, as a highly efficient and high throughput approach, the wide use of chemical proteomics in drug target identification has enhanced our confidence in improving our understanding of the molecular mechanisms of these drugs. With the aid of chemical proteomics, an unprecedented number of biological targets have been tested, and various technologies emerging today provide us with a superior platform to further investigate drug targets (Huang, *et al.*, 2012).

Proteomics has unique and significant advantages as an important complement to a genomics approach in the complex drug-discovery process. It provides unbiased, large-scale protein analysis at the cost of being a relatively low-throughput technique. Proteomic approaches have been applied to the process of drug discovery and development with different purposes, such as biomarker discovery, target identification and characterization (Schirle *et al.*, 2012; Colzani and Carini, 2014).

1.7 Proteomics

The term proteomics describes the study and characterization of complete set of proteins present in a cell, organ, or organism at a given time. In general, proteomic approaches can be used
(a) for proteome profiling,

(b) for comparative expression analysis of two or more protein samples,
(c) for the localization and identification of posttranslational modifications, and
(d) for the study of protein–protein interactions (Chandramouli and Qia, 2009).

The emergence of proteomics, the large-scale analysis of proteins, has been inspired by the realization that the final product of a gene is inherently more complex and closer to function than the gene itself. Shortfalls in the ability of bioinformatics to predict both the existence and function of genes have also illustrated the need for protein analysis. Moreover, only through the study of proteins can posttranslational modifications be determined, which can profoundly affect protein function. Proteomics has been enabled by the accumulation of both DNA and protein sequence databases, improvements in mass spectrometry, and the development of computer algorithms for database searching (Chandramouli and Qia, 2009).

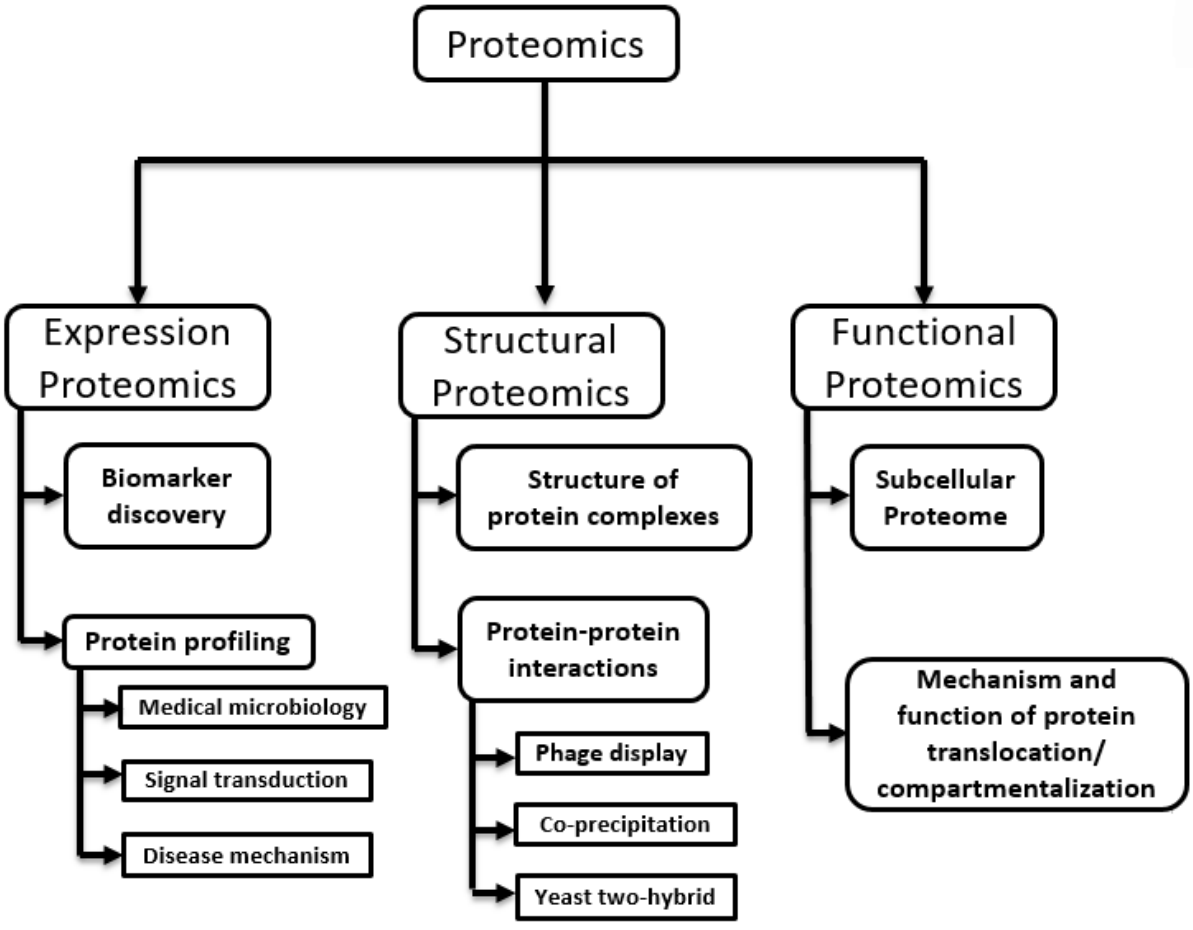


Figure 10: Types of proteomics and their applications to biology (Graves and Haystead, 2002; Lau *et al.*, 2003).

In the quest to characterize the proteome of a given cell or organism, it should be remembered that the proteome is dynamic. The proteome of a cell will reflect the immediate environment in which it is studied. In response to internal or external cues, proteins can be modified by posttranslational modifications, undergo translocations within the cell, or be synthesized or degraded. Thus, examination of the proteome of a cell is like taking a "snapshot" of the protein environment at any given time. Considering all the possibilities, it is likely that any given genome can potentially give rise to an infinite number of proteomes (Graves and Haystead, 2002).

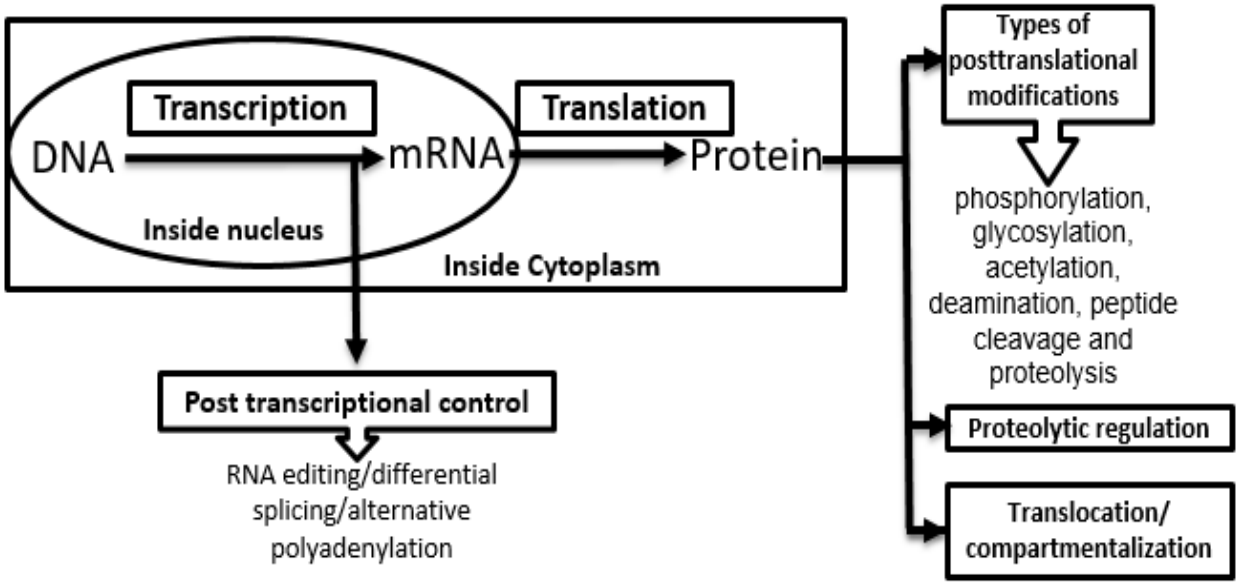


Figure 11: Gene and protein expression regulation or modification from transcription to post translation (Lau *et al.*, 2003; Banks *et al.*, 2000).

1.8 The development of proteomics: a brief history

The idea of observing the protein expression of genomes in a holistic manner rather than one protein at a time arose with the advent of 2-D gels and the concept of the proteome itself was introduced by Marc Wilkins in 1994 at a conference in Siena, Italy, having coined the term earlier that year in association with his then PhD supervisor Keith Williams. The words 'proteome' and 'proteomics' have been widely adopted by the biological community. In the 10 years since their introduction, their use has grown very rapidly.

Initially, proteomics researchers had a goal of visualizing all proteins from a proteome on a single, or perhaps one acidic range and one basic range (2-D) polyacrylamide gel. This was

happening in the late 1980s, and there was enormous excitement about the possibility of being able to see all proteins in a proteome. However, it did not take long to realize that the separation and visualization of all proteins from a proteome was not a straightforward task. In the mid-1990s, the availability of the first genome sequences and predicted proteomes allowed theoretical 2-D gels to be calculated, showing where each protein spot should be found. This revealed a bimodal distribution of proteins, with the majority of proteins having isoelectric point (pI) 4–6.5 and another group of proteins having pI 8–12. Most proteins had a mass of less than 100 kDa. The comparison of these theoretical maps with experimental 2-D gel separations immediately highlighted shortcomings with 2-D gels in that they were poor in resolving very acidic, very basic or very high mass proteins. A meta-analysis of proteins seen on 2-D gels and those predicted theoretically from genomes of *Escherichia coli*, *Saccharomyces cerevisiae* and *Bacillus subtilis* highlighted two additional issues (Thongboonkerd, 2007). The first was that hydrophobic proteins were largely absent from the 2-D gels and that low-abundance proteins present at less than 1,000 copies per cell were likely to be undetectable, owing to limitations on the loading capacity and staining sensitivity of the 2-D gel process.

Since that time, a series of important technical advances have been made to help us see more proteins in the proteome. Broadly speaking, a number of strategies have been adopted. These include the running of narrow pI range gels to ‘zoom in’ on a particular region of the proteome, the fractionation of samples into either biologically (e.g. organelles) or physicochemically distinct fractions (e.g. membrane proteins) that can then be analyzed appropriately, the enrichment or depletion of proteins of interest from a sample, along with new solubilisation and gel running techniques to assist in the analysis of the more difficult proteins. Importantly, fractionation has provided an avenue to load more of the relevant portion of samples of interest onto 2-D gels, thus assisting in the detection of lower-abundance proteins.

To completely bypass many of the challenges of working with complex mixtures of proteins, a conceptually different strategy emerged for protein analysis in proteomics. Called ‘shotgun proteomics’, probably inspired by the shotgun DNA sequencing approaches that were developed by Venter *et al.* (1998), it involves taking complex mixtures of proteins or indeed a whole proteome, and digesting all proteins to peptides with endoproteinases of known specificity. The resulting mixtures of peptides, which are physicochemically more homogenous than their parent

proteins although greater in number, are then analyzed using 2-D liquid chromatography and tandem mass spectrometry. Peptide fragment data are matched against sequence databases to determine the proteins present in a sample. Whilst this approach has limitations, it provides an alternative to gel based analyses for the separation and identification of large numbers of proteins from a proteome (Moritz *et al*., 2004).

Chapter Two
PROTEOMIC TECHNOLOGIES

2.1 Common proteomic technologies

Proteins are the principal targets of drug discovery. Most large pharmaceutical companies now have a proteomics-oriented biotech or academic partner or have started their own proteomics division. Common applications of proteomics in the drug industry include target identification and validation, identification of efficacy and toxicity biomarkers from readily accessible biological fluids, and investigations into mechanisms of drug action or toxicity. Target identification and validation involves identifying proteins whose expression levels or activities change in disease states. These proteins may serve as potential therapeutic targets or may be used to classify patients for clinical trials. Proteomics technologies may also help identify protein-protein interactions that influence either the disease state or the proposed therapy. Efficacy biomarkers are used to assess whether target modulation has occurred. They are used for the characterization of disease models and to assess the effects and mechanism of action of lead candidates in animal models. Toxicity (safety) biomarkers are used to screen compounds in pre-clinical studies for target organ toxicities as well as later on in development during clinical trials (Walgren and Thompson, 2004).

Table 3: Common proteomic technologies, applications, and their limitations (Chandramouli and Qia, 2009; Meyfour *et al.*, 2013).

Technology	Application	Strengths	Limitations
Two-Dimensional Electrophoresis (2DE)	Protein separation; Quantitative expression profiling	Relative quantitative; Post-translational modifications (PTM) information.	Poor separation of acidic, basic, hydrophobic and low abundant proteins; difficulty in detection of scarce proteins; less reproducibility; and incompatibility for high molecular weight, or high pI protein analysis.
Difference Gel Electrophoresis	Protein separation; Quantitative	Relative quantitative; Post-translational	Proteins without lysine cannot be labeled;

Technology	Application	Strengths	Limitations
(DIGE)	expression profiling	modifications information; Reduction of intergel variability; High sensitivity.	requires special equipment for visualization and fluorophores are very expensive.
Isotope-Coded Affinity Tag (ICAT)	Chemical isotope labeling for quantitative proteomics	Sensitive and reproducible; Detect peptides with low expression levels.	Selective detection of proteins with high cysteine residues and acidic proteins are not detected; expensive method.
Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC)	Direct isotope labeling of cells; Differential expression pattern	Degree of labelling is very high; Quantitation is straightforward	Cannot be applied to tissue protein analysis or labeling of tissue samples
Isobaric Tag for Relative and Absolute Quantitation (iTRAQ)	Isobaric tagging of peptides	Multiplex several samples; Relative quantification High-throughput	Increases sample complexity; require fractionation of peptides before MS.
Multidimensional Protein Identification Technology (MUDPIT)	Identification of protein-protein interactions; Deconvolve complex sets of proteins	High separation; Large protein complexes identification	Not quantitative; difficulty in analyzing the huge data set; difficult to identify isoforms
Protein array	Quantitate specific proteins used in diagnostics	High-throughput; Highly sensitive Low sample	Limited protein production; poor expression methods; availability of the antibodies;

Technology	Application	Strengths	Limitations
	(biomarkers or antibody detection) and discovery research	consumption	accessing very large numbers of affinity reagents.
Mass spectrometry	Primary tool for protein identification and characterization	High sensitivity and specificity; High-throughput; Qualitative and quantitative PTM information	No individual method to identify all proteins. Not sensitive enough to identify minor or weak spots. MALDI and ESI do not favor identification of hydrophobic peptides and basic peptides.
Bioinformatics	Analysis of qualitative and quantitative proteomic data	Knowledge and discovery from mass spectrometric data; functional analysis, data mining, and	No integrated pipeline for processing and analysis of complex data. Search engines do not yield identical results.

2.2 Gel electrophoresis

One way to identify proteins is to extract all the proteins from a sample of cells and separate them in a gel matrix, using a technique called polyacrylamide gel electrophoresis (PAGE). The proteins are separated by size, with the smaller proteins moving faster through the gel than the larger proteins. After staining, a pattern of bands appears that corresponds to the proteins in the cell. However, this technique can only resolve a few hundred proteins, and cannot separate proteins of very similar size (Lodish *et al.*, 2000).

A modification of this procedure - called 2D gel electrophoresis - separates proteins into two dimensions, using two different characteristics. Proteins are separated in the first dimension by their isoelectric point (pI), the specific point at which the net charge of the protein is zero. These separated proteins, in a flat gel strip, are then placed on a standard polyacrylamide gel. Every protein band that was separated in the first dimension according to its isoelectric point is now

separated in the second dimension by its size. The result is small spots, each representing a protein; even proteins of the same size will be resolved if they have a different isoelectric point. A good 2D gel can resolve one thousand to two thousand proteins, which appear, after staining, as dots in the gel. This technique is useful when comparing two similar samples to find specific protein differences; for example, comparing the proteins in a tumor cell versus a normal cell. However, it can miss very small proteins or non-abundant proteins (Lodish *et al.*, 2000).

Two-dimensional electrophoresis of proteins has preceded, and accompanied, the birth of proteomics (Rabillouda and Lelong, 2011). The basic concept of 2D electrophoresis is schematized in the following diagram.

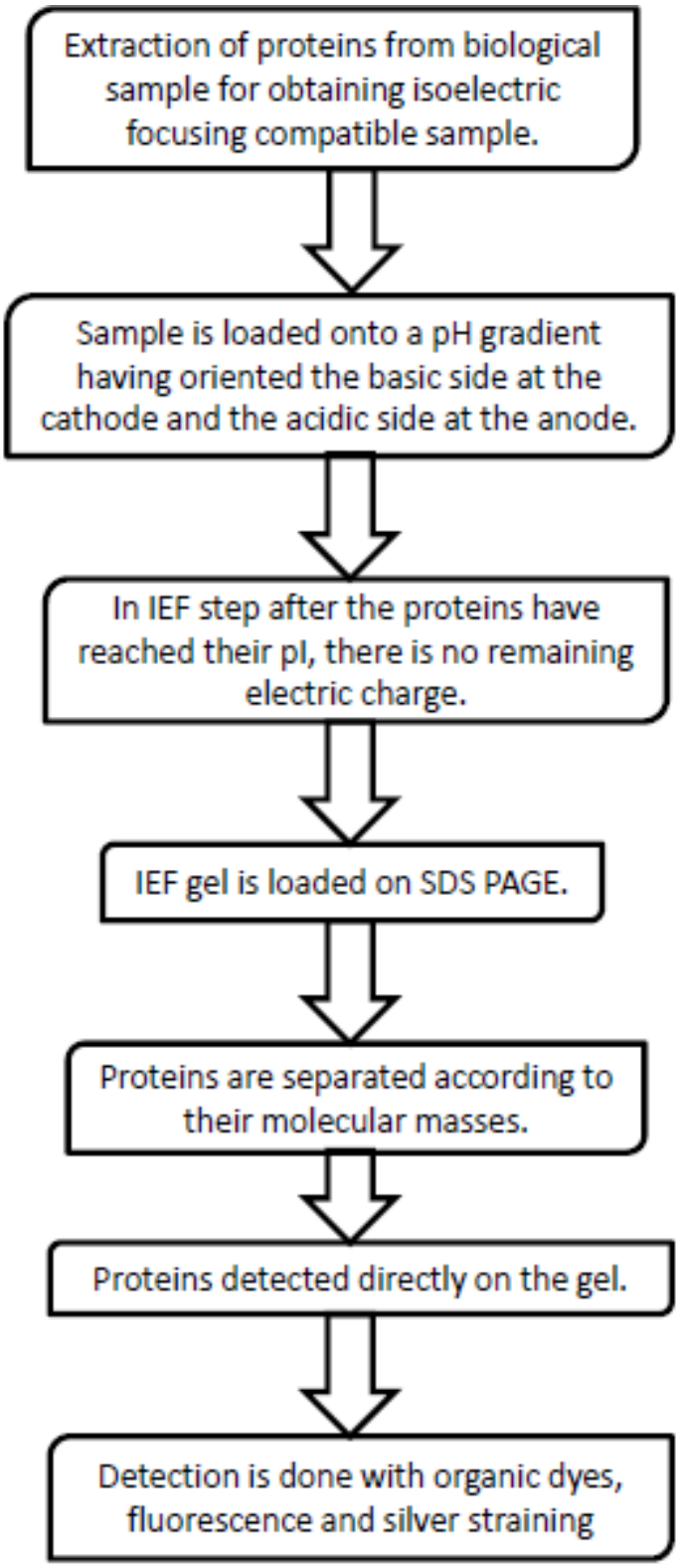


Figure 12: Scheme of principle of 2D gel electrophoresis (Rabillouda and Lelong, 2011).

The total process starts with the extraction of proteins from the biological sample to get an isoelectric focusing (IEF) compatible sample. The sample is then loaded onto a pH gradient oriented with the acidic side at the anode and the basic side at the cathode. After the IEF step, the proteins have reached their pI and thus have no remaining electrical charge. The strip is then equilibrated in a sodium dodecyl sulfate (SDS) containing buffer, so that all proteins becomes strongly negatively charged. The IEF gel is then loaded on top of a SDS PAGE gel, and the proteins are separated according to their molecular masses. After this step, the proteins are detected directly on the gel. Current in-gel protein detection methods fall into three major categories: detection with organic dyes, silver staining and fluorescence (Rabillouda and Lelong, 2011).

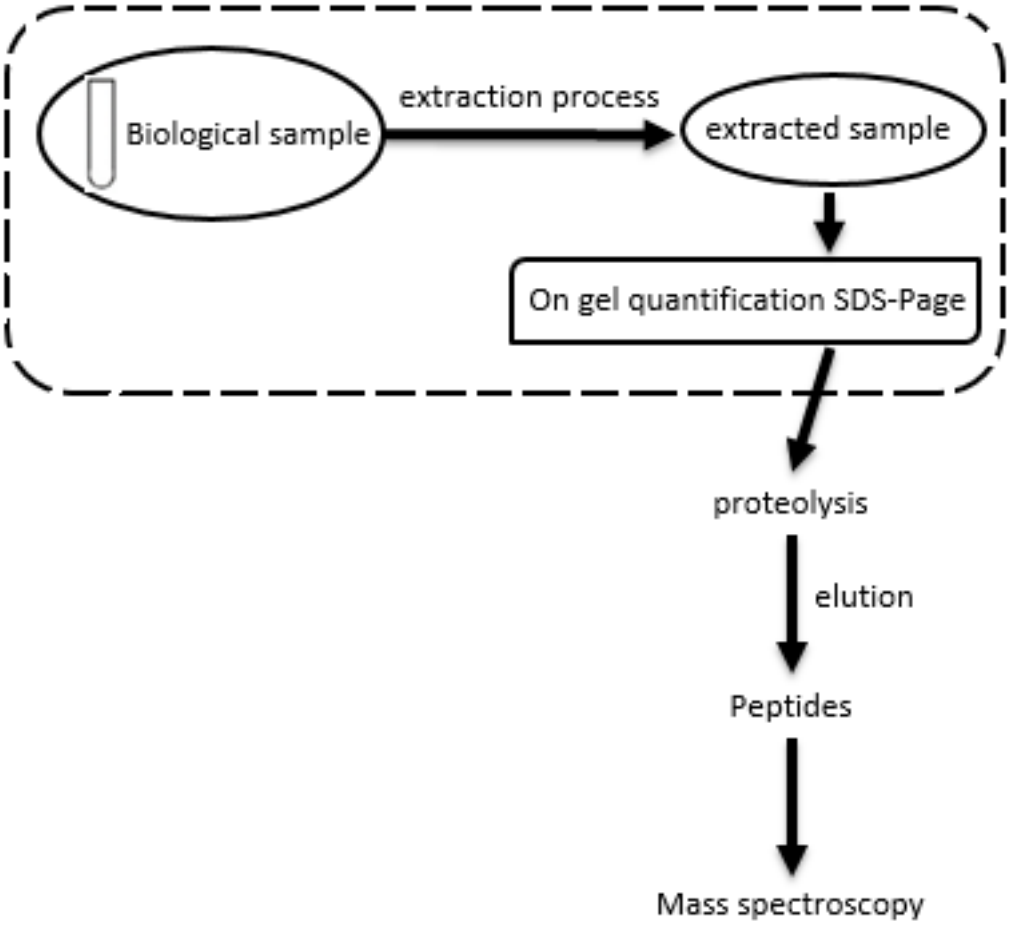


Figure 13: Scheme of principle of 2D gel-based proteomics (Rabillouda and Lelong, 2011).

The first steps are exactly those described on Figure 4. Then after detection of the proteins in the gels, the resulting images are quantitatively analyzed to determine the spots of interest. Those spots are then excised and submitted to in-gel digestion (generally with trypsin). The resulting peptides are then eluted and analyzed by mass spectrometry (MS), leading to protein identification and characterization. The dotted box shows the part of 2D electrophoresis in the whole process, and it can be easily seen that key steps, including sample preparation and quantitative analysis, take place during this process (Rabillouda and Lelong, 2011).

Two-dimensional gel electrophoresis (2DGE) is a widely used separation method in which proteins are first separated by isoelectric focusing (IEF) and then by molecular mass in the second dimension on application of an electrical potential across a solid-based gel. One of the recent technical advances in 2DGE has been the development of the multiplexing fluorescent two-dimensional fluorescence difference gel electrophoresis (2D-DIGE) method, which relies on direct labeling of the lysine groups on proteins with cyanine (Cy) dyes before IEF. A critical aspect of the use of 2D-DIGE technology is the ability to label 2–3 samples with different dyes and electrophorese all the samples on the same 2D gel, thus reducing spot pattern variability and the number of gels in an experiment and yielding simple and accurate spot matching (Tannu and Hemby, 2006).

2.3 Isotope-Coded Affinity Tag (ICAT)

Gel-free, or MS based, proteomics techniques are emerging as the methods of choice for quantitatively comparing proteins levels among biological proteomes, since they are more sensitive and reproducible than two-dimensional gel-based methods. ICAT is one of the most employed chemical isotope labeling methods and the first quantitative proteomic method to be based solely on using MS. ICAT is a cysteine specific, protein-based labeling strategy designed to compare two different sample states (Gygi *et al.*, 1999; Chandramouli and Qia, 2009).

Each ICAT reagent consists of three essential groups: a thiol-reactive group, an isotope-coded light or heavy linker, and a biotin segment to facilitate peptide enrichment. In an ICAT experiment, protein samples are first labeled with either light or heavy ICAT reagents on cysteine thiols. One sample is labeled with a light isotope and the other with a heavy isotope, and then the samples are combined. The mixtures of labeled proteins are then digested by trypsin and

separated through a multistep chromatographic separation procedure. Peptides are identified with tandem MS and the relative quantifications of peptides are inferred from the integrated liquid chromatograph (LC) peak areas of the heavy and light versions of the ICAT-labeled peptides (Chandramouli and Qia, 2009).

The ratios of signal intensities of the ICAT-tagged peptide pairs are quantified to determine the relative levels of proteins in the two samples. The specificity of ICAT reagents for cysteine residues means that the approach is sometimes preferred because it reduces sample complexity. However, this also creates a drawback in that peptides lacking cysteine residues will not be labeled, so many important peptides, including those with post-translational modifications (PTMs) will be discarded (Gygi *et al*., 1999).

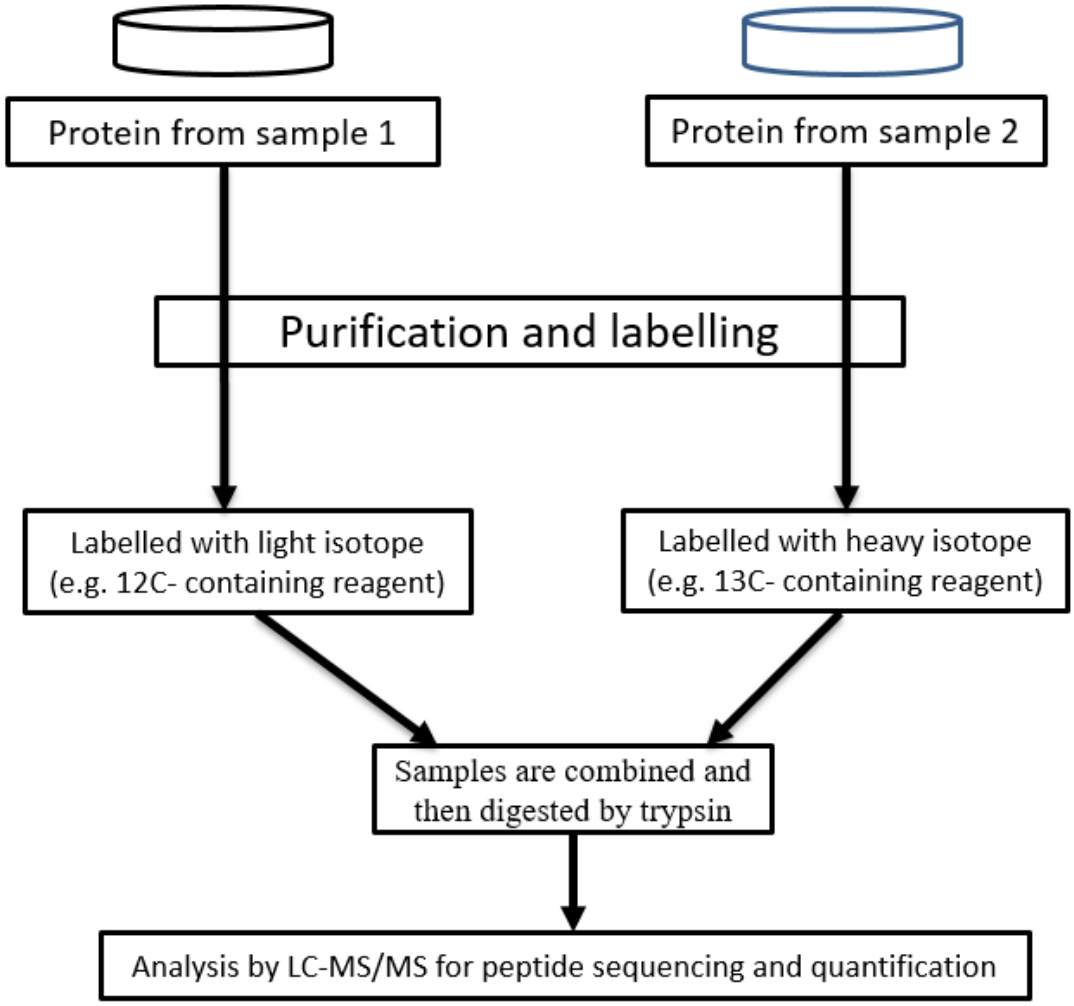


Figure 14: Isotope-Coded Affinity Tags (ICAT) (Gygi et al., 1999; Ross *et al.*, 2004).

2.4 Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC)

Stable Isotope Labeling by Amino acids in Cell culture (SILAC) is a simple, robust, yet powerful approach in mass spectrometry (MS)-based quantitative proteomics. SILAC labels cellular proteomes through normal metabolic processes, incorporating non-radioactive, stable isotope containing amino acids in newly synthesized proteins. Growth medium is prepared where natural ('light') amino acids are replaced by 'heavy' SILAC amino acids. Cells grown in this medium incorporate the heavy amino acids after five cell doublings and SILAC amino acids have no effect on cell morphology or growth rates. When light and heavy cell populations are mixed, they remain distinguishable by MS and protein abundances are determined from the relative MS signal intensities (Ong and Mann, 2006).

Co-precipitating interacting proteins with an epitope tagged protein as the affinity bait is a popular approach to characterize a protein's functional role in the cell. Classically performed with western blotting, applying proteomics instead provides an unbiased and sensitive readout of proteins interacting directly with the bait or in a larger protein complex. This proteomics experiment can identify co-precipitating proteins from a specific biological pathway, providing enormous insight into the functional role of the bait molecule. We have successfully applied proteins, small-molecules and nucleic acids as affinity baits in proteomics (Blagoev *et al*., 2003; Ong *et al*., 2002).

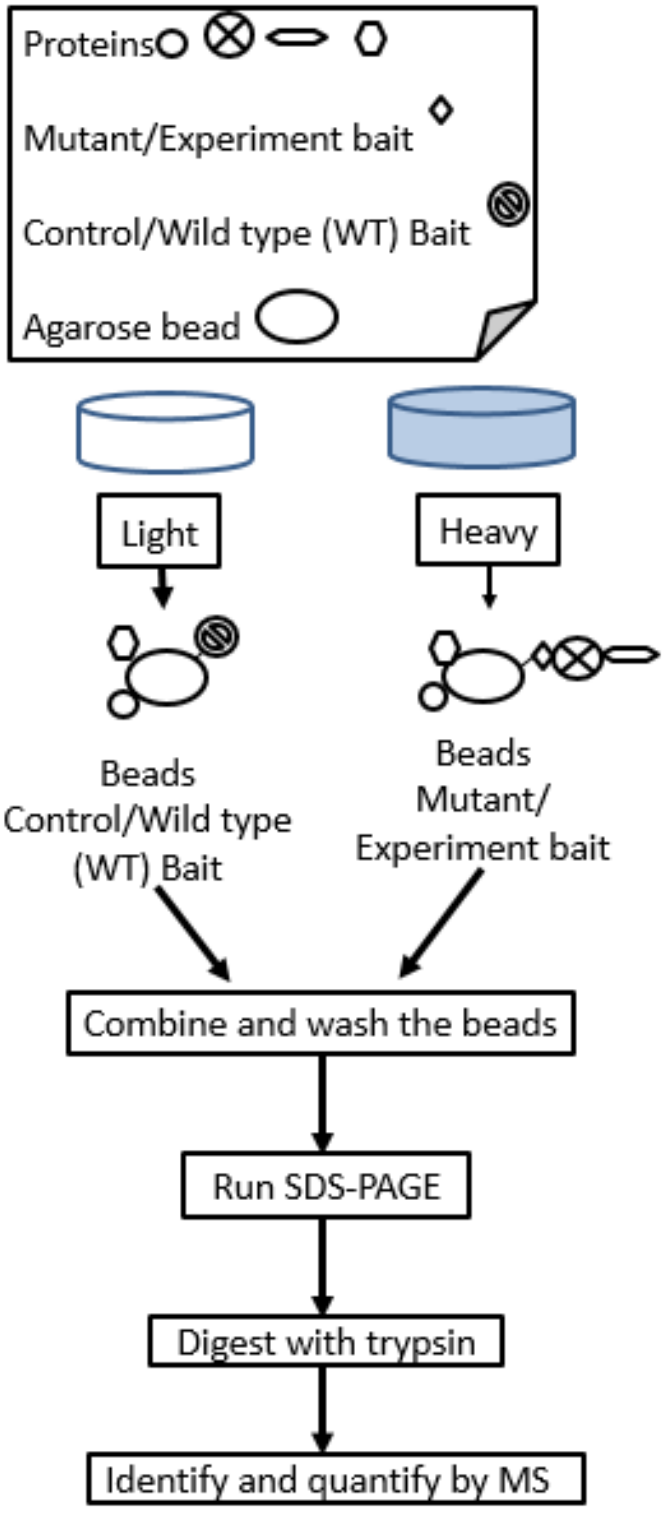


Figure 15: Co-precipitating interacting proteins with an epitope tagged protein as the affinity bait (Blagoev et al., 2003; Ong et al., 2002).

2.5 Isobaric Tag for Relative and Absolute Quantitation (iTRAQ)

The iTRAQ reagent is well known for relative and absolute quantitation of proteins. The iTRAQ technology offers several advantages, which include the ability to multiplex several samples, quantification, simplified analysis and increased analytical precision and accuracy. The interest of this multiplexing reagent is that 4 or 8 analysis samples can be quantified simultaneously. In this technique, the introduction of stable isotopes using iTRAQ reagents occurs on the level of proteolytic peptides. This technology uses an ester derivative to modify primary amino groups by linking a mass balance group (carbonyl group) and a reporter group (based on N-methylpiperazine) to proteolytic peptides via the formation of an amide bond. Due to the isobaric mass design of the iTRAQ reagents, differentially labeled peptides appear as a single peak in MS scans, reducing the probability of peak overlapping. When iTRAQ-tagged peptides are subjected to MS/MS analysis, the mass balancing carbonyl moiety is released as a neutral fragment, liberating the isotope-encoded reporter ions which provides relative quantitative information on proteins. An inherent drawback of the reported iTRAQ technology is due to the enzymatic digestion of proteins prior to labelling, which artificially increases sample complexity and this approach needs a powerful multidimensional fractionation method of peptides before MS identification (Chandramouli and Qia, 2009).

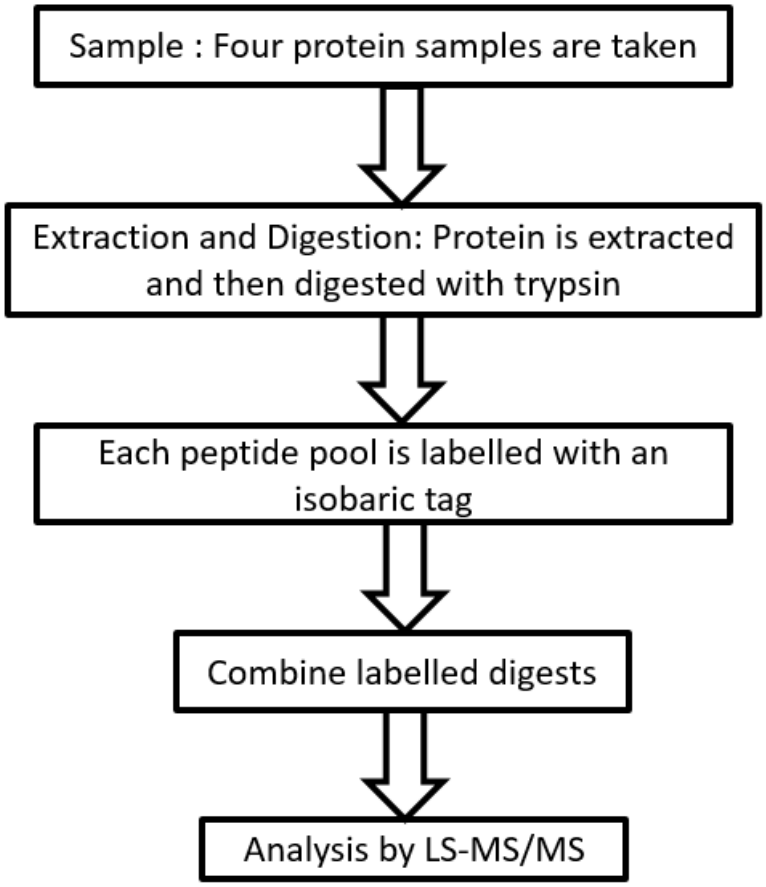


Figure 16: iTRAQ work flow (Chandramouli and Qia, 2009).

2.6 Multidimensional Protein Identification Technology (MUDPIT)

Another approach to analyse proteomes without gels is "shotgun" analysis using MudPIT. In the MudPIT approach, protein samples are subject to sequence specific enzymatic digestion, usually with trypsin and endoproteinase lysC, and the resultant peptide mixtures are separated by strong cation exchange (SCX) and reversed phase (RP) high performance liquid chromatography (HPLC). Peptides from the RP column enter the mass spectrometer and MS data is used to search the protein databases. The MudPIT technique generates an exhaustive list of proteins present in a particular protein sample, it is fast and sensitive with good reproducibility however, it lacks the ability to provide quantitative information (Chandramouli and Qia, 2009).

2.7 Protein microarray

Another strategy for the large-scale study of proteins is similar to the DNA microarrays, which measure gene expression in different cells types. Based on the rapid, large-scale technology (often called high-throughput technology) that was developed for DNA microarrays, scientists have developed similar microarrays for proteins. In a protein microarray, very small amounts of different purified proteins are placed on a glass slide in a pattern of columns and rows. These proteins must be pure, fairly concentrated, and folded in their active state. Various types of probe molecules may be added to the array and assayed for ability to bind or react with the protein. Typically the probe molecules are labeled with a fluorescent dye, so that when the probe binds to the protein it results in a fluorescent signal that can be read by a laser scanner (Annenberg Learner, 2015).

This technology can complement other techniques, such as mass spectrometry and yeast two-hybrid assays, to identify thousands of protein-protein interactions. Protein arrays can be screened for their ability to bind other proteins in a complex, receptors, antibodies; lipids; enzymes; peptides; hormones; specific DNA sequences; or small molecules, such as potential new drugs. One of the most promising applications for protein microarrays is the rapid detection or diagnosis of disease by identifying a set of proteins associated with the disease (Annenberg Learner, 2015).

One example of the use of this technique is the development of a microarray that may help in the treatment of cancer. This microarray contains many different mutant forms of a protein called p53. P53 is an anti-cancer protein, called a "tumor-suppressor protein," and about half of all cancers have mutations in p53. Researchers can screen the immobilized mutant p53 proteins in the microarray for biological activity, as well as for new drugs that can restore its normal tumor-suppressing function (Annenberg Learner, 2015).

2.8 Mass spectrometry

Regardless of the choice of a given proteomic separation technique, gel-based or gel-free, a mass spectrometer is always the primary tool for protein identification. During the last decade, significant improvements have been made in the application of MS for the determination of

proteinsequences. Mass spectrometers consist of an ion source, the mass analyzer, and an ion detection system. Analysis of proteins by MS occurs in three major steps

(a) protein ionization and generation of gas-phase ions,

(b) separation of ions according to their mass to charge ratio, and

(c) detection of ions (Chandramouli and Qia, 2009).

In gel-free approaches such as ICAT and MudPIT, samples are directly analyzed by MS whereas, in gel-based proteomics (2DE and 2D-DIGE), the protein spots are first excised from the gel and then digested with trypsin. The resulting peptides are then separated by LC or directly analyzed by MS. The experimentally derived peptide masses are correlated with the peptide fingerprints of known proteins in the databases using search engines (e.g., Mascot, Sequest) (Chandramouli and Qia, 2009).

There are two main ionization sources which include matrix assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) and four major mass analyzers, which are time-of-flight (TOF), ion trap, quadrupole, and fourier transform ion cyclotron (FTIC) which are currently in use for protein identification and characterization. A combination of different mass analyzers in tandem such as quadrupole-TOF and quadrupole-ion trap has combined the individual strengths of different types of mass analyzers and greatly improved their capabilities for proteome analysis (Chandramouli and Qia, 2009).

Simple mass spectrometers such as MALDI-TOF are used for only measurement of mass, whereas tandem mass spectrometers are used for amino acid sequence determination. Tandem mass spectrometry or MS/MS is performed by combining two different MS separation principles. In tandem MS, individual trypsin-digested peptides are fragmented after a liquid phase separation (Chandramouli and Qia, 2009).

2.9 Mass Spectrometry to Identify Cellular Proteins

While the 2D gel method easily separates proteins, it doesn't identify them. If there are differences in spots between the proteins in a cancer cell and a normal cell, this method cannot determine the actual identity of the different proteins in the two cell types. To identify these proteins, individual spots are excised from 2D gels and then subjected to mass spectrometry, which separates charged particles, or ions, according to mass. First the molecules in the sample

are ionized to produce a population of charged molecules. A mass analyzer then separates the sample's molecules based on their mass to charge ratio. A detector then produces a peak for each ion; this peak gives the mass and represents the amount of the ion. A computer program reads the complex spectral information from the mass spectrometry process. The program matches the information on the each peptide's mass against the mass of theoretical, predicted peptides, based on known proteins in databases. This is called peptide mass mapping. With many different peptides for each protein, the computer can match the sequence to one or more known proteins. Peptide mass mapping can only be used in situations where the genome has been sequenced and all predicted proteins for the genome are known (Huang *et al.*, 2012).

Another application of mass spectrometry is protein fingerprinting. This technique has been used to identify unique sets of proteins in blood, which serve as markers for different forms of cancer. Interestingly, for this method to be useful, we do not need to know the actual identities of the particular proteins used as markers for a disease. Instead, this technique relies on pattern recognition software. Using training data from samples from individuals with and without cancer, the program searches for a particular pattern of peaks that correlates with cancer. This technique requires only a drop of blood and does not require any detailed genetic information; however, its accuracy in predicting some forms of cancer is limited because the number of marker peptides is not sufficiently large. As more samples are evaluated, the accuracy will likely increase because the software will be able to find more accurate peptide patterns correlating to cancer. Proteomic fingerprinting holds great promise as a diagnostic tool for a variety of diseases that produce distinctive patterns of proteins in blood (Huang *et al.*, 2012).

2.10 Global proteomic approaches

While the majority of existing approaches aim to discover new protein targets for a specific drug type, novel drug target discovery also has other uses in which the proteins identified might be involved in a certain disease or pathologic state. One notable advantage of global proteomic profiling methods over other methods lies in that proteomic approaches require no purification and are unbiased because the unmodified drug interacts with its endogenous targets. Based on this rationale, global proteomic methods have been widely used in the pharmaceutic industry. For example, MudPIT is typically used to gain more knowledge regarding the mechanisms of action between natural products and targets involved in the maintenance of a particular

phenotype. The MS analysis method mentioned above constitutes a powerful and universal method for such unbiased studies (Huang, *et al.*, 2012).

Currently, it is possible to explore the interactions between drugs and their targets using MS and protein microarray techniques. Other quantitative proteomic approaches could be used for the identification of novel drug targets, including methods involving differential labeling with stable isotopes, such as ICAT or iTRAQ. These techniques are usually used to simultaneously quantify alterations in protein abundance. In addition, another approach named DIGE can detect changes in protein expression levels, and these proteins can likewise be labeled with fluorescent probes. For example, the expression of urinary proteins was analyzed using DIGE and MALDI-MS. This study group identified meprin1α as a potential drug target for sepsis induced acute renal failure (ARF). However, the DIGE method could hardly detect low abundance proteins. In addition, membrane proteins cannot be easily analyzed by this approach. The histones isolated from individual biological replicates can be digested to their tryptic peptides, followed by LC-FT-MS/MS analysis during different MS workflow protocols. In addition, techniques such as SILAC and 2DE are also important tools in global proteomic approaches (Huang, *et al.*, 2012).

Chapter Three

PROTEOMICS AND DRUG DISCOVERY

3.1 New drug target

Proteome related information is used to identify the proteins associated with the disease. Then by a computer software that protein will be used as a target for new drug (Verrills, 2006).

- Certain protein implicated in the disease is identified.
- Three dimensional protein structure is generated using software.
- Drugs are designed using computer programme to interfere with the actions of the protein for example when a molecule fits the active site of an enzyme and cannot be released then it will inactivate proteins (Neha and Harikumar, 2013).

3.2 Chemical proteomics

Many biochemical and cell biological methods are used to study the mode of action for new drug molecules. Chemical proteomics is one of the new and widely used methods. This technique uses small, drug-like molecules either exposed to protein chips or bound to a resin. Proteins binding the ligand are then considered as potential drug targets. Other than understanding the selectivity and mechanism of action of a compound, chemical proteomics can also identify previously unknown protein targets for a compound that may help in the identification of potential side effects. In proteomics, specific chemical probes are used to survey the activities of certain classes of proteins. These precisely designed probes can be used for proteomics, mainly in protein expression analysis and identification as well as cellular localization and regulation (Neha and Harikumar, 2013).

In the last few years, the application of probes has become a main focus of pharmaceutic companies for the development of novel drug. Trifunctional probes are used to identify several protein targets involved in some crucial cell processes. For example, cysteine proteases have been found to play an important role in malarial infection, cataract formation, and apoptosis. Selective probes are now being designed to target this group of proteins (Huang, *et al*., 2012).

Table 4: List of drug target identified using chemical proteomics (Huang, *et al*., 2012; Wissing *et al.*, 2004; Wuemail *et al.*, 2015).

Drug	Involved targets	Chemical proteomics	Disease
Pyrido[2,3-d] pyrimidine	Src, PDGFR, FGFR, RICK, p38α	Affinity chromatography, nano-HPLC MS/MS, LC-MS/MS	Cancer
SB 203580	RICK, CK1, GAK, PKNβ, JAK1	Affinity chromatography, MS	Inflammatory diseases
Imatinib	BCR-ABL, ABL, c-KIT, PDGFR, NQO2, c-fms	LC-ESI-MS/MS, HPLC-MS, ELISA, western blotting	Chronic myeloid leukemia
Nilotinib	c-ABL, BCR-ABL, c-KIT, PDGFR, ARG NQO2, DDR1	LC-ESI-MS/MS, HPLC-MS, immunoblotting	Chronic myeloid leukemia
Dasatinib	c-ABL, BCR-ABL, BCR-ABL, DDR1, BTK, TEC	LC-ESI-MS/MS, immunoblotting, SDS/PAGE, LC-MS/MS	Chronic myeloid leukemia
Bosutinib	ABL and SRC family kinases, STE and TEC family kinases, CAMK2G	Affinity chromatography, MS, kinobeads/iTRAQ	Chronic myeloid leukemia
(R)-Roscovitine	CDKs, PDXK	Affinity chromatography, electrophoresis, and western blotting	Cancers, neurodegenerative diseases, viral infections, and glomerulonephritis
GF109203X	PKC, Ste20-related kinase, adenosine	SDS-PAGE separation, MS,	Cancers, heart failure

Drug	Involved targets	Chemical proteomics	Disease
	kinase, quinine reductase type 2, voltage-dependent sodium channels, and the 5-HT 3 receptor	immunoprecipitation	
Bisindolylmaleimide-III	PKC-α, GSK3-β, CaMKII, adenosine kinase, CDK2, quinine reductase type 2, PKAC-R, prohibitin, VDAC, and heme binding proteins	Mass spectrometry, affinity chromatography	Cancers, heart failure
SU6668	β-PDGFR, VEGFR2, FGFR, Yes and Lyn, RSK3, AMPKα1, and ULK3	16-BAC/SDS-PAGE, MS, immunofluorescence	Cancer
Thalidomide	bFGF, CRBN	Immunofluorescence, immunoblotting, and ELISA	Myeloma, erythema nodosumleprosum, and leprosy
Arsenic trioxide	CDK6, cdc2, cyclin A, PML-RARa and PML	Western blotting, immunoprecipitation	Acute promyelocytic leukemia

Abbreviations: AMPKα1, the AMP-activated protein kinase α1; bFGF, basic fibroblast growth factor; BTK, Bruton's tyrosine kinase; CAMK2G, calcium/calmodulin-dependent protein kinase type II gamma chain; CDKs, cyclin-dependent kinases; CK1, Casein kinase 1; CRBN, cereblon; DDR1, discoidin domain receptor-1; FGFR, fibroblast growth factor receptor; GAK, cyclin G-associated kinase; JAK1, Janus kinase 1; NQO2, quinone oxidoreductase 2; PDGFR, platelet-

derived growth factor receptor; PDXK, pyridoxal kinase; PKC, protein kinase C; PKNβ, protein kinase N beta; RICK, Rip-like interacting caspase-like apoptosis-regulatory protein kinase; VDAC, voltage-dependent anion channel; ULK3, Unc-51-like kinase 3.

3.3 Binding sites discovered by proteomic approach

Protein kinases are crucial for controlling the elements of cellular signaling, which is de-regulated in several diseases for example in human cancer. So the protein kinase enzyme family has become a major class of drug targets in recent years. Pharmacological inhibition of protein kinases can be achieved with small molecule inhibitors, which block the catalytic activity of kinases by interfering with ATP binding (Wissing *et al.*, 2004).

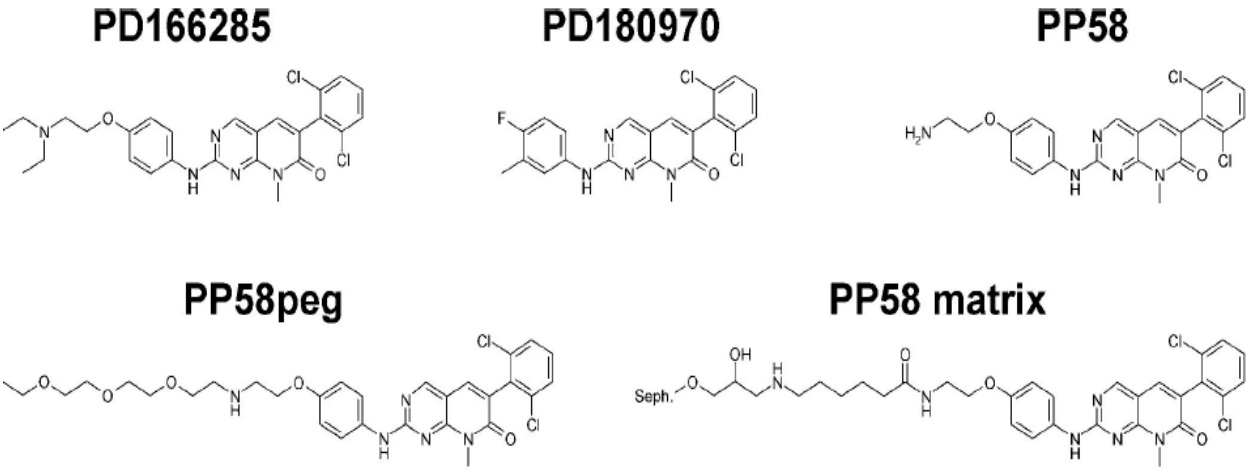


Figure 17: Chemical structures of pyrido[2,3-d]pyrimidine-based compounds (Wissing *et al.*, 2004).

Pyrido[2,3-d]pyrimidines were initially developed as broadly active inhibitors of several tyrosine kinases such as Src, platelet-derived growth factor receptor and fibroblast growth factor receptor (FGFR). The pyrido[2,3-d]pyrimidine PP58 is related in structure to PD166285 and PD180970. Pegylation of PP58 yielded PP58peg (Wissing *et al.*, 2004).

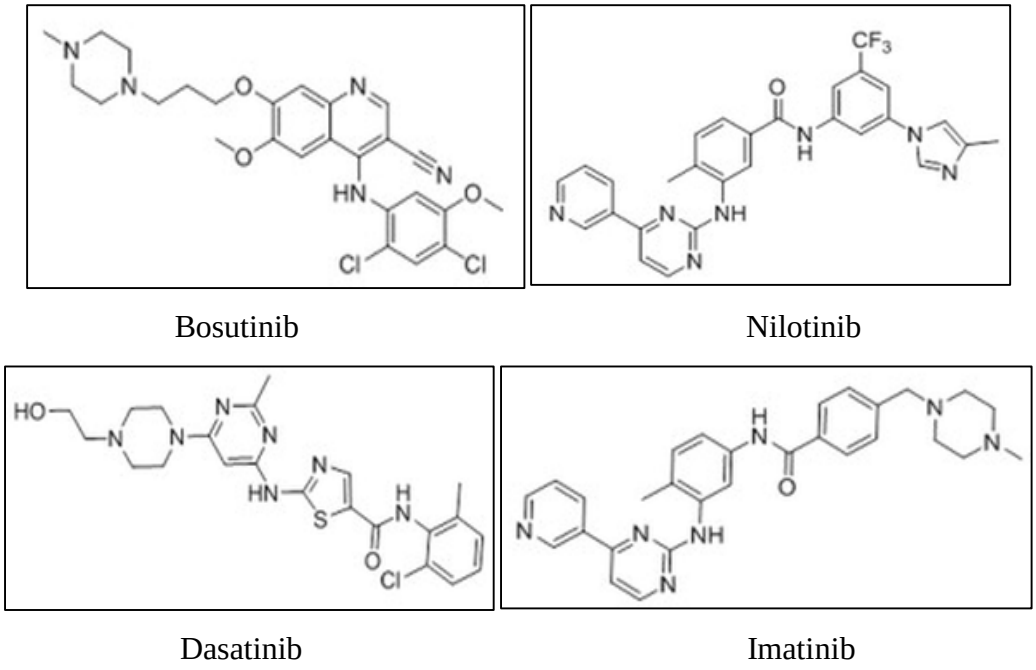


Figure 18: FDA-approved small-molecule kinase inhibitors (Wuemail *et al.*, 2015).

Protein kinases represent an important class of drug targets, particularly in oncology and inflammation. However, kinase drug discovery epitomizes the shortcomings of the single-gene/single-protein/single-assay paradigm, as kinase inhibitors can be both conformation specific and have multiple targets, as demonstrated by recently launched multi-kinase drugs. Evidently, compounds directed at the ATP-binding site of kinases are not likely to be specific for a single kinase, because humans have around 500 protein kinases and more than 2,000 other purine-binding proteins that share similar binding pockets (Bantscheff *et al.*, 2007).

Recent progress in affinity-based proteomic strategies has enabled the direct determination of protein-binding profiles of small-molecule drugs under more physiological conditions. To date, methods rely on the attachment of labels to the compound (immobilization, fluorescent or affinity tags) or to the proteins, which may introduce artifacts driven by the altered properties of the compound or the protein. Here we describe a chemical proteomics methodology that

i. captures a large portion of the expressed kinome and related bproteins on a mixed kinase-inhibitor matrix (kinobeads) and
ii. subsequently analyzes this defined subproteome by quantitative mass spectrometry (MS) (Bantscheff *et al.*, 2007).

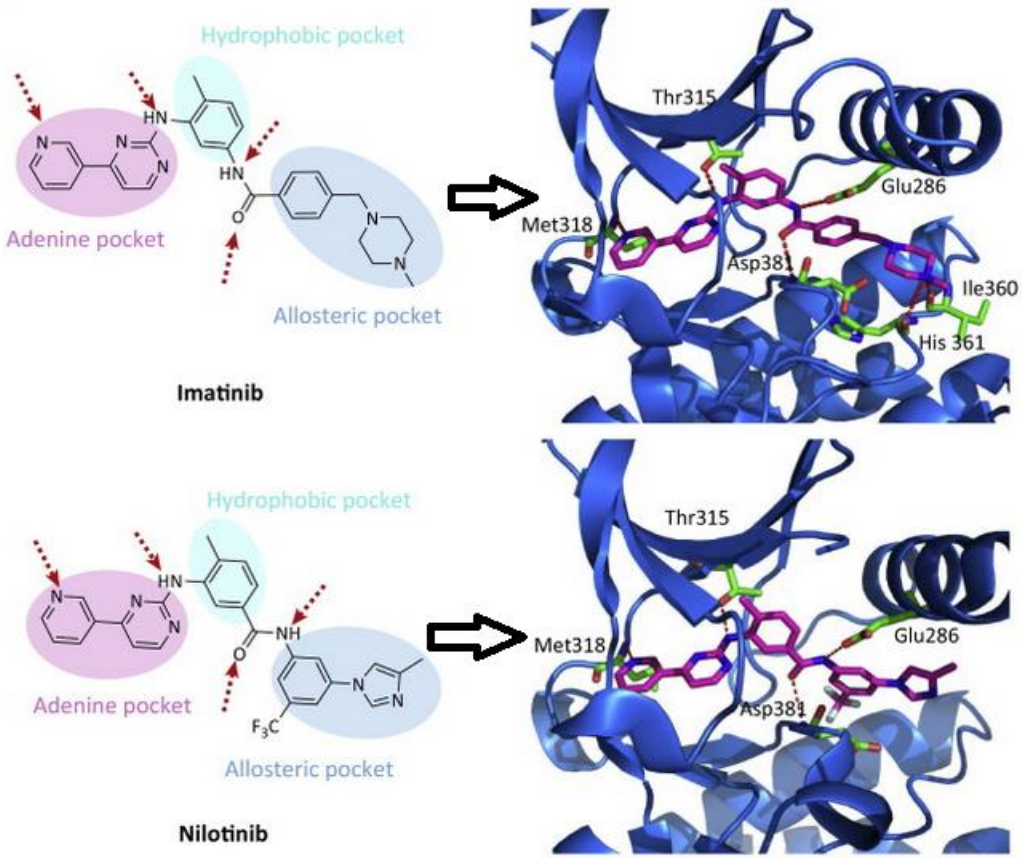


Figure 19: Binding sites of Nilotinib and Imatinib (Wuemail *et al*., 2015).

3.4 Use of proteomics to identify biomarkers

Biomarkers are biological molecules found in blood, tissues, or other body fluids that are a sign of a condition or disease, or of a normal or abnormal process. They may also be used to see how well the body responds to a treatment for a disease or condition. The accessibility of disease-related proteins in tissues and bodily fluids has triggered extensive protein-focused research for the hunt of "biomarkers." Proteomics can find out the protein contents of a variety of biospecimens and accurately measure the concentrations of these proteins. This can provide scientists and clinicians with the knowledge to understand the various processes involved in disease development and progression and hopefully help to identify biomarkers specific for these cellular processes along with those indicating efficacious therapeutic intervention (Yocum *et al.*, 2006; Swan *et al.*, 2013).

Table 5: List of biomarkers identified for diagnosis of several diseases (Chen *et al.*, 2004; Meuwis *et al.*, 2007; Ornstein *et al.*, 2006).

Disease	Clinical Biomarker
Alzheimer's disease	Cystatin C and peptic fragment of the neurosecretory protein; Tau proteins in CSF; protein kinase C in red blood cells
Multiple sclerosis	CSF cystatin C and matrix metalloproteinases in serum
Metastasic breast cancer	Protein kinase C
Traumatic brain injury	Hyperphosphorylated axonal neuro-filment protein
Gefitinib resistance	Hypoxia-inducible factor-1 in head and neck cancer, epithelial membrane protein-1
Stroke	Lipoprotein associated phospholipase-A2, nucleoside diphosphate kinase-A, intracellular adhesion molecule 1
Ischemic heart disease	Creatine kinase, fatty acid binding protein, troponin, myoglobin and natriuretic peptide
Congestive heart failure	G protein-coupled receptor kinase-2
Inflammatory bowl disease	Human platelet antigen 2, Platelet Factor 4
Gliomas	Receptor protein tyrosine phosphatase-B
Artherosclerotic heart disease	Lipid-phospholipase-A2, lipid-modified proteins and adipocyte-enhancer binding protein
Prostate cancer	Annexin

Table 6: List of biomarkers identified for drug development (Yocum *et al.*, 2006; Jhanker *et al.*, 2012; Swan *et al.*, 2013).

Biomarker	Drugs	Molecular Target	Cancer Type
Estrogen receptor (IHC)	Tamoxifen, aromatase inhibitors	ER	Breast Cancer
HER2 gene amplification (IHC, FISH)	Trastumab, lapatinib, neratinib, pertuzumab, trastuzumab-DM1	HER2 receptor	Breast and upper GI cancers
BCR-ABL translocation	Imatinib, dasatinib, nilotinib	ABL kinase	CML
EGFR kinase domain mutations (not T790M)	Erlotinib, gerfitinib	EGFR kinase	NSCLC
PML-RAR translocation	All-trans retinoic acid	PML-RAR	APL
BRCA1/2 mutation	Olaparib, veliparib	PARP	Breast and ovarian cancer
B-RAF V600E mutation	Vemurafenib	B-RAF kinase	Melanoma
Mutant KIT	Imatinib	KIT kinase	GIST
Mutant PDGFRβ	Imatinib	PDGFR kinase	CMML
EML4/ALK translocation	crizotinib	ALK kinase	NSCLC
EGFR-T790M	Afatinib + cetuximab	EGFR (drug resistant)	NSCLC
*TSC1*mutations	Everolimus	Torc1/2	Pancreas NET
PIK3CA hot spot mutations	P13K inhibitors	P110α	Breast, endometrial, colon cancer

Abbrevations: APL-Acute Promyelocytic Leukemia; CML-Chronic Myologenous Leukemia; CMML- Chronic Myolomonocytic Leukemia; GI-Gastrointestinal; GIST-Gastrointestinal Stromal Tumour; PDGFR-Platelet-Derived Growth Factor receptor.

There are several sources of biomarker, although blood is the most-used biomarker discovery matrix to date (Collings and Vaidya, 2008; Omenn, 2006). Other bio-specimens are also in use to overcome the shortcomings of profiling proteins in blood. These alternative specimens include:

- Tumor biopsy tissue (Sitek *et al.*, 2005),
- Cancer cell lines such as cathepsin D for prostate (Sardana *et al.*, 2007) and colon cancer (Volmer *et al.*, 2005),
- Soluble-secreted proteins and shed membrane proteins from tumour cells (Ahn and Simpson, 2007),
- Saliva for cancer, auto immune disease (Seibert *et al.*, 2005),
- Bile for biliary malignancy (Bonney *et al.*, 2008),
- Ventricular CSF for Neurologic disease (Choe *et al.*, 2006),
- Urine (Collings and Vaidya, 2008; Kelloff and Sigman, 2005).

Table 7: Some of the techniques used for biomarker development (Collings and Vaidya, 2008; Sardana *et al.*, 2007; Jhanker *et al.*, 2012).

Technology	Method	Objective	Tissue
Genomics	SNP genotyping	Identify susceptibility or disease modifying gene	Nucleated cells, diseased tissue
	Positional cloning/microsatellites	Fine mapping/sequencing of disease loci	
	Expression analyses	Identification of signaling pathways and differential expression of genes	

Technology	Method	Objective	Tissue
Proteomics	2DGE, MS, LC-MS, GC-MS, MS-MS, MALDI-TOF MS	Identification of low-abundance proteins, their subcellular location, posttranslational modification, interactions among proteins	Urine, blood, saliva, tissues
Bioinformatics	BLAST, hierarchical clustering, SOM	Link microarray data to biological pathways	Data from various techniques
Pharmacogenetics	SNP genotyping	Relate genetic makeup to drug response	Nucleated cells

Abbreviations: BLAST indicates basic local alignment search tool; GC-MS, gas chromatography–mass spectrometry; LC-MS, liquid chromatography–mass spectrometry; MALDI-TOF, matrix-assisted laser desorption–ionization time-of-flight mass spectrometry; MS, mass spectrometry; MS-MS, tandem mass spectrometry; SNP, single nucleotide polymorphism; SOM, self-organizing map; and 2DGE, 2-dimensional gel electrophoresis.

Proteome profiling is a method of biomarker discovery that generated great interest among the scientists (Bonney *et al.*, 2008; Seibert *et al.*, 2005) in the recent years. Analyzing the proteome content of blood or several other body fluids over the course of disease progression could reveal potential biomarkers indicative of specific disease status that may be used extensively in future medical diagnostics.

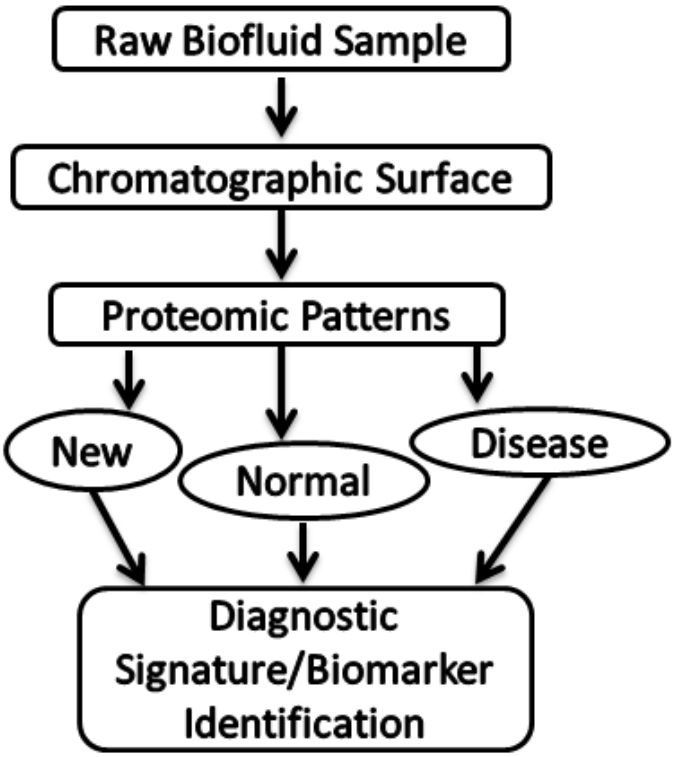


Figure 20: Proteome profiling using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF/MS) (Veenstra, 2007).

The process of proteome profiling can be briefly described as below (Bonney *et al.*, 2008; Veenstra, 2007):

1. Samples are selected to address the clinical question.
2. Proteins separated from raw bio-fluid sample by applying to a chip made up of a specific chromatographic surface. Proteins are allowed to bind to the surface, which is then washed to remove non-binding species.
3. The mass spectra of the several hundreds proteins bound to the chip spot are then recorded using a simple time-offlight mass spectrometer from disease-affected patients and healthy controls.
4. Data are analysed bioinformatically to identify potential differentially expressed proteins or biomarkers compared to the healthy controls and classify the samples as coming from diseased, healthy or from an unknown condition.
5. Suitable assays are developed, results are validated.

This method of proteome profiling is a high-throughput process and has a great advantage of capability of analyzing and comparing hundreds of bio-fluid samples in a matter of days. Many individual studies showed great results in the ability to correctly classify the sources of bio-fluid samples from either healthy or diseased individuals (Veenstra, 2007).

3.5 Identification and assignment of candidate target

Drug targets are proteins or signal transduction pathways in which proteins are involved. Therapeutic relevancy of the chosen target must be proven first prior to initiating any other processes in drug discovery. Some diseases and potential targets are listed below.

Table 8: Some disease conditions with corresponding targets (Katayama *et al.*, 2007; Kopec *et al.*, 2005).

Disease	Potential Target
Cancer	Maleate dehydrogenase (Primary target)
Tumour vascularization	Tyrosin kinase receptor, arora kinases and TANK-binding kinase-1
Malaria	Aldehyde dehydrogenase-1 and quinine reductase-2
Inflammation	RICK (Rip-like interacting kinase), CLARP (caspase-like apoptosis-regulatory protein kinase) and GAK (cyclin-G associated kinase)

Recombinant Protein Microarray and Computational Drug Design are the two unique techniques serving the purpose of identifying drug targets, target validation and 3D structure elucidation upon which a new drug molecule is being searched against the chosen target that usually involves high-throughput screening, wherein large libraries of chemicals are tested to determine their ability to modify the target (Meuwis *et al.*, 2007).

3.6 Toxicity determination

Proteins whose levels are modified in response to drug administration could provide vital clues with respect to drug effectiveness and toxicity. These proteins will serve as efficacy or toxicity

biomarkers to guide clinical trial studies. Similarly, analyses of protein profiles before and after pharmacological treatments could also confirm the mechanism of drug action and provide insight for new drug discovery. A biomarker consortium was launched in 2006, in response to FDA's Critical Path Initiative published in March 2004 reinforcing the requirement for additional biomarkers to predict drug toxicity in preclinical studies, specifically biomarkers that can act as surrogate endpoints and/or aid in making efficacious and cost-saving decisions or terminating drug development more quickly (Collings *et al*., 2008). Kim *et al.*, 2008 have been successfully implicated proteomics to understand the effects of TCDD (2,3,7,8-tetrachlorodibenzo-p-dioxin), a toxic environmental pollutant and a potent liver carcinogen in various diseases. To understand the mechanisms of TCDD toxicities they analysed of the glycoproteins and phosphoproteins in Chang cells. Using 2-DE and MS, several candidate biomarkers that are potentially involved TCDD toxicities were identified. There are several other examples of toxicity biomarkers that are given in the table below.

Table 9: Biomarkers for the diagnosis of the effect of chemicals (Sinha *et al.*, 2007; Collings *et al*., 2008).

Toxicants	Biomarkers
Aminophenol and D-serine, rodent nephrotoxins	Fumarylacetoacetate hydrolase, a toxicity-associated plasma protein
Automobile emission and waste incineration	Transthyretin, sarcolectin and haptoglobin
Ethylene glycol monomethyl ether, cyclophosphamide, sulfasalazine and 2,5-hexanedione	Glutathione S-transferase, glyceraldehyde phosphate dehydrogenase, phosphotidylethanolamine-binding protein
Hydrazine exposure	Glucose and lipid metabolizing enzymes; oxidative stress related proteins
Neurotoxins	Glial fibrillary acid protein
Skeletal muscle toxicity	Urinary parvalbumin-alpha
Steasis causing hepatotoxins	Pyruvate dehydrogenase, phenylalanine hydroxylase, chaperone-like protein, glucose-regulated protein, serum paraoxonase, serum albumin, and peroxiredoxin

3.7 Pharmacoproteomics

Pharmacoproteomics is the use of proteomic technologies in drug discovery and development. Pharmacoproteomics, along with pharmacogenomics and pharmacogenetics, play an important role in the development of personalized medicines in several ways. Proteomic technologies are contributing to molecular diagnostics, which is a basis of personalized medicine. Pharmacoproteomics is a more functional representation of patient-to-patient variation. Proteomics-based characterization of multifactorial diseases may help to match a particular target-based therapy to a particular marker in a subgroup of patients. Individualized therapy may be based on differential protein expression rather than a genetic polymorphism. Finally, proteomic technologies allow discovery and development of drugs suitable for personalized therapy. In clinical diagnostics protein chips are increasingly used, mainly in the point-of-care diagnostics. This will allow the practice of personalized medicine in the clinic in the future (Jain, 2004).

3.8 Glycoproteomics

Glycosylation is a structurally complicated and diverse type of protein modifications. Protein glycosylation plays crucial functions in many biological processes, including disease genesis and progression. Glycoproteomics is the classification of proteins modified by carbohydrates. Glycoproteomic studies usually include methods to enrich glycoproteins having particular carbohydrate structures from protein mixtures and followed by quantitative proteomic analysis. These glycoproteomic studies evaluate which proteins are glycosylated, the glycosylation sites, the carbohydrate structures, and also the abundance and role of the glycoproteins in different biological and pathological processes (Wong, 2005).

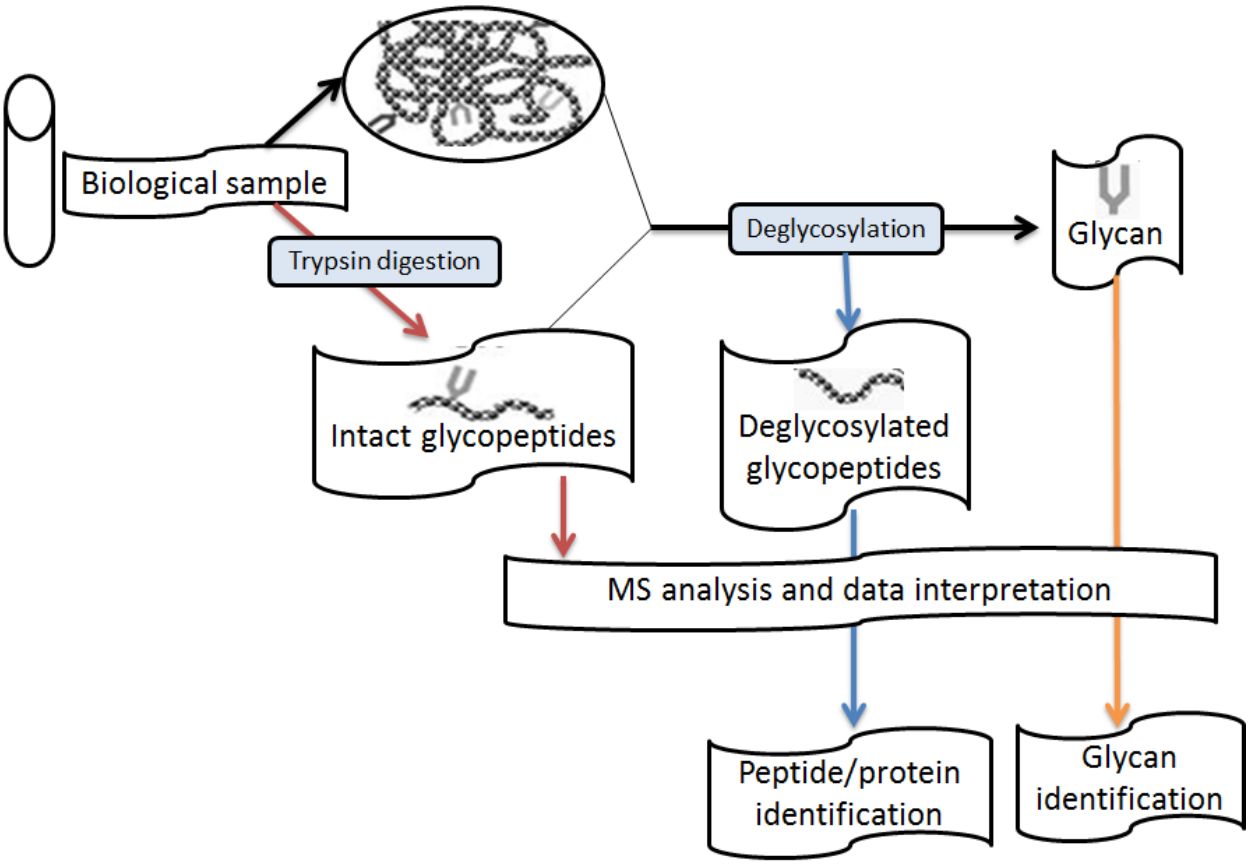


Figure 21: Glycoproteomic analysis using mass spectrometry based strategies (Bunkenborg *et al.*, 2004; Xin *et al.*, 2008).

Potential glycosylation sites may and may not be occupied by different glycans and different forms have different biological functions or result in clinical abnormalities. There are several reasons that glycoproteins are of clinical interests:

(i) Protein glycosylation influences properties of the proteins and also regulates various biological functions through particular protein–carbohydrate recognition.

(ii) The abnormal glycosylation of glycoproteins is a fundamental characteristic of disease genesis and progression

(iii) Extra-cellular proteins are usually glycosylated and can enter the blood stream. These proteins can provide biomarkers for certain disease states.

(iv) Glycoproteomic analysis greatly decreases sample complexity and significantly increases the detection sensitivity for low abundance proteins (Wong, 2005).

Table 10: List of some of the FDA approved cancer biomarkers (Bunkenborg *et al.*, 2004; Xin *et al.*, 2008).

Protein target	Source	Detection	Clinical biomarker	Glycosy-lation	Disease
α-Fetoprotein	Serum	Glycoprotein	Diagnosis	Yes	Nonseminomatous testicular cancer
Human chorionic gonadotropin-β	Serum	Glycoprotein	Diagnosis	Yes	Testicular cancer
CA19–9	Serum	Carbohydrate	Monitoring	Yes	Pancreatic cancer
CA125	Serum	Glycoprotein	Monitoring	Yes	Ovarian cancer
CEA (carcinoembryonic antigen)	Serum	Protein	Monitoring	Yes	Colon cancer
Epidermal growth factor receptor	Tissue	Protein	Therapy selection	Yes	Colon cancer
KIT	Tissue	Protein (IHC)	Diagnosis/ Therapy selection	Yes	Gastrointestinal (GIST) cancer
Thyroglobulin	Serum	Protein	Monitoring	Yes	Thyroid cancer
PSA-prostate-specific antigen (Kallikrein 3)	Serum	Protein	Screening/ Monitoring /Diagnosis	Yes	Prostate cancer

Protein target	Source	Detection	Clinical biomarker	Glycosy-lation	Disease
CA15–3	Serum	Glycoprotein	Monitoring	Yes	Breast cancer
CA27–29	Serum	Glycoprotein	Monitoring	Yes	Breast cancer
HER2/NEU	Tissue, Serum	Protein (IHC), Protein	Prognosis/ Therapy selection/M onitoring	Yes	Breast cancer
Fibrin/FDP-fibrin degradation protein	Urine	Protein	Monitoring	Yes	Bladder cancer
BTA-bladder tumour-associated antigen (Complement factor H related protein)	Urine	Protein	Monitoring	Yes	Bladder cancer
CEA and mucin (high molecular weight)	Urine	Protein (Immunofluor escence)	Monitoring	Yes	Bladder cancer

Chapter Four
CONCLUSION

4.1 Conclusion

The drug discovery is not a predefined process. Modern approaches include target-based drug discovery where researchers survey the proteins. For this type of technology, the two most important requirements are to find more effective biomarkers for disease detection and to discover proteins that can be targeted with therapeutic drugs. As many proteins have multiple functions, it is important to develop drugs for each function of a protein. In addition, most proteins act as a part of networks and complexes and this may also affect the way function of a protein in a cell. This may also influence the ability of drugs to disable the protein. With the knowledge about the proteome, the structure and function of each protein, and the protein-protein interactions and their complexities will be critical for developing the most effective diagnostic techniques and disease treatments in the future (Jhanker *et al.*, 2012).

Chapter Five
REFERENCE

Reference

Ahn, S. M. and Simpson, R. J. (2007) Body fluid proteomics: prospects for biomarker discovery. *Proteomics-Clinical Application* 1, 1004-1015.

Alberts, B., Johnson, A., Lewis, J., Raff, M. Roberts, K. and Walter, P. (2002) *Molecular Biology of the Cell*. 4th edn, New York, Garland Science.

Annenberg Learner (2015) Unit: 2-Proteins and proteomics: Protein Microarrays. *Annenberg Foundation* [Online] Available from: http://www.learner.org/courses/biology/textbook/proteo/proteo_11.html [Accessed 3rd October 2015]

Atlas of Genetics and Cytogenetics in Oncology and Haematology (2015) *Amino Acids (reminder).* [Online] Available from: http://atlasgeneticsoncology.org/Educ/AminoAcidsEngID30066ES.html [Accessed 3rd October 2015].

Banks, R. E., Dunn, M. J., Hochstrasser, D. F., Sanchez, J. C., Blackstock, W., Pappin, D. J. and Selby, P. J. (2000) Proteomics: new perspectives, new biomedical opportunities. *Lancet* 356 (9243), 1749-1756.

Bantscheff, M., Eberhard, D., Abraham, Y., Bastuck, S., Boesche, M., Hobson, S., Mathieson, T., Perrin, J., Raida, M., Rau, C., Reader, V., Sweetman, G., Bauer, A., Bouwmeester, T., Hopf, C., Kruse, U., Neubauer, G., Ramsden, N., Rick, J., Kuster, B. and Drewes, G. (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol.* 25 (9), 1035-1044.

Berg, J. M., Tymoczko, J. L. and Stryer, L. (2002) *Biochemistry*. 5th edn, New York, W. H. Freeman.

Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M., Foster, L. J. and Mann, M. (2003) A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat Biotechnol* 21 (3), 315-318.

Bonney, G. K., Craven, R. A., Prasad, R., Melcher, A. F., Selby, P. J. and Banks, R. E. (2008) Circulating markers of biliary malignancy: opportunities in proteomics? *The Lancet Oncology* 9, 149-158.

Branden, C. I. and Tooze, J. (1998) *Introduction to Protein Structure*. 2nd edn, New York, Garland Science.

Bunkenborg, J., Pilch, B. J., Podtelejnikov, A. V. and Wisniewski, J. R. (2004) Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* 4, 454–465.

Chandramouli, K. and Qia, P. Y. (2009) Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Hum Genomics Proteomics* 2009.

Chen, Y., Chen, G. and Xia, W. (2004) Identification of hypoxia-regulated proteins in head and neck cancer by proteomic and tissue array profiling. *Cancer Research* 64 (20), 7302-7310.

Choe, L. H., Werner, B. G. and Lee, K. H. (2006) Two-dimensional Protein Electrophoresis: From Molecular Pathway Discovery to Biomarker Discovery in Neurological Disorders. *NeuroRX* 3 (3), 327-335.

Collings, F. B. and Vaidya, V. S. (2008) Novel technologies for the discovery and quantitation of biomarkers of toxicity. *Toxicology* 245, 167-174.

Colorado State University (2002) *Amino Acids.* [Online] Available from: http://www.vivo.colostate.edu/hbooks/molecules/aminoacids.html [Accessed 3rd October 2015].

Colzani M. and Carini M. (2014) Proteomics-based drug discovery and chemoproteomics. *Applications of High-Resolution Mass Spectrometry in Drug Discovery and Development* 120-139.

Graves, P. R. and Haystead, T. A. J. (2002) Molecular Biologist's Guide to Proteomics. *Microbiol Mol Biol Rev.* 66 (1), 39–63.

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* 17 (10), 994-999

Huang, F., Zhang, B., Zhou, S., Zhao, X., Bian, C. and Wei, Y. (2012) Chemical proteomics: terra incognita for novel drug target profiling. *Chinese Journal of Cancer* 31 (11), 507-518.

Jain, K. K. (2004) Role of pharmacoproteomics in the development of personalized medicine. *Pharmacogenomics* 5 (3), 331-336.

Jhanker, Y. M., Kadir, M. F., Khan, R. I., and Hasan, R. (2012) Proteomics in Drug Discovery. *Journal of Applied Pharmaceutical Science* 2 (8), 01-12

Kelloff, G. J. & Sigman, C. C. (2005) New science-based endpoints to accelerate oncology drug development. *European Journal of Cancer* 41 (4), 491-501.

Kim, J. H., In, Y. J., Kim, W. K., Bae, K. H., Kang, S. and Lee, S. C. (2008) Differential signatures of protein glycosylation and phosphorylation in human Chang liver cells induced by TCDD treatment. *Toxicology Letters* 178 (1), 20-28.

Kopec, K. K., Bozyczko-Coyne, D. and Williams, M. (2005) Target identification and validation in drug discovery: the role of proteomics. *Biochemical Pharmacology* 69 (8), 1133-1139.

Lau, A. T., He, Q. Y., Chiu, J. F. (2003) Proteomic technology and its biomedical applications. *Acta biochimica et biophysica sinica* 35 (11), 965-975.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D. and Darnell, J. (2000) *Molecular Cell Biology*. 4th edition, New York, W. H. Freeman and Company.

Meuwis, M. A., Fillet, M., Geurts, P., Seny, D., Lutteri, L., Chapelle, J. P., Bours, V., Wehenkel, L., Belaiche, J., Malaise, M., Louis, E. & Merville, M. P. (2007) Biomarker discovery for inflammatory bowel disease, using proteomic serum profiling. *Biochemical pharmacology* 73 (9), 1422-1433

Meyfour, A., Tavirani, M. R., Sadeghi, M. R. (2013) Common Proteomic Technologies, Applications, and their Limitations. *Journal of Paramedical Sciences* 4, 15-125.

Moritz, R. L., Ji, H., Schütz, F., Connolly, L. M., Kapp, E. A., Speed, T. P. and Simpson, R. J. (2004) A proteome strategy for fractionating proteins and peptides using continuous free-flow electrophoresis coupled off-line to reversed-phase high-performance liquid chromatography. *Analytical chemistry* 76 (16), 4811-4824.

Neha, S. and Harikumar, S. L. (2013) Use of genomics and proteomics in pharmaceutical drug discovery and development: A review. *International Journal of Pharmacy and Pharmaceutical Science* 5 (3), 24-28.

Nelson, D. L. and Cox, M. M. (2008) *Lehninger Principles of Biochemistry.* 5th edn, New York, W. H. Freeman.

Omenn, G. S. (2006) Strategies for plasma proteomic profiling of cancers. *Proteomics*, 6, 5662-5673.

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1(5), 376-86.

Ong, S. E. and Mann, M. (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc* 1(6), 2650-2660

Ornstein, D. K. and Tyson, D. R. (2006) Proteomics for theidentification of new prostate cancer biomarkers. *Urol Oncol.* 24 (3), 231-236.

Rabillouda, T. and Lelong, C. (2011) Two-dimensional gel electrophoresis in proteomics: A tutorial. *Journal of Proteomics* 74 (10), 1829–1841.

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovsk,i N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3 (12), 1154-1169.

Sardana, G., Marshall, J. and Diamandis, E. P. (2007) Discovery of candidate tumor markers for prostate cancer via proteomic analysis of cell culture-conditioned medium. *Clinical Chemistry* 53, 429-437.

Schirle, M., Bantscheff, M. and Kuster, B. (2012) Mass spectrometry-based proteomics in preclinical drug discovery. *Chem. Biol.* 19 (1), 72 – 84

Seibert, V., Ebert, M. P. and Buschmann, T. (2005) Advances in clinical cancer proteomics: SELDI-ToF-mass spectrometry and biomarker discovery. *Briefings in Functional Genomics and Proteomics*, 4, 16-26.

Sitek, B., Luttges, J., Marcus, K., Kloppel, G., Schmiegel, W., Meyer, H. E., Hahn, S. A. and Stuhler, K. (2005) Application of fluorescence difference gel electrophoresis saturation labelling for the analysis of microdissected precursor lesions of pancreatic ductal adenocarcinoma. *Proteomics* 5, 2665-2679.

Swan, A. L., Mobasheri, A., Allaway, D., Liddell, S. and Bacardit, J. (2013) Application of Machine Learning to Proteomics Data: Classification and Biomarker Identification in Postgenomics Biology. *OMICS*. 17 (12), 595–610.

Tannu, N. S. and Hemby, S. E. (2006) Two-dimensional fluorescence difference gel electrophoresis for comparative proteomics profiling. *Nat Protoc*. 1 (4), 1732–1742.

Thongboonkerd, V. (2007) *Proteomics of Human Bodyfluids: Principles, Methods, and Applications*. New York, Humana Press.

Veenstra, T. D. (2007) Global and targeted quantitative proteomics for biomarker discovery. *Journal of Chromatography B.* 847, 3-11.

Venter, J. C., Adams, M. D., G. G., Sutton, Kerlavage, A. R., Smith, H. O. and Hunkapiller, M. (1998) Shotgun Sequencing of the Human Genome. *Science* 280 (5369), 1540-1542.

Verrills, N. M. (2006) Clinical Proteomics: Present and Future Prospects. *Clin Biochem Rev.* 27 (2), 99–116.

Volmer, M. W., Stuhler, K., Zapatka, M., Schoneck, A., Klein-Scory, S., Schmiegel, W., Meyer, H. E. and Schwarte-Waldhoff, I. (2005) Differential proteome analysis of conditioned media to detect Smad4 regulated secreted biomarkers in colon cancer. *Proteomics* 5, 2587-2601.

Walgren, J. L. and Thompson, D. C. (2004) Application of proteomic technologies in the drug development process. *Toxicol Lett.* 149 (1-3), 377-385.

Wissing, J., Godl, K., Brehmer, D., Blencke, S., Weber, M., Habenberger, P., Stein-Gerlach, M., Missio, A., Cotton, M., Müller, S. and Daub, H. (2004) Chemical proteomic analysis reveals alternative modes of action for pyrido[2,3-d]pyrimidine kinase inhibitors. *Mol Cell Proteomics* 3 (12), 1181-1193.

Wong, C. H. (2005) Protein glycosylation: new challenges and opportunities. *J. Org. Chem.* 70, 4219–4225

Wuemail, P., Nielsen, T. E. and Clausen, M. H. (2015) FDA-approved small-molecule kinase inhibitors. *Trends in Pharmacological Sciences* 36 (7), 422–439.

Wu, G. (2013) *Amino Acids: Biochemistry and Nutrition*. CRC Press, Boca Raton.

Xin, L., Li, M. and Jianjun, L. (2008) Recent developments in the enrichment of glycopeptides for glycoproteomics. *Anal. Lett*. 41, 10.

Yocum, A. K., Busch, C. M., Felix, C. A., Blair, I. A. (2006) Proteomics-based strategy to identify biomarkers and pharmacological targets in leukemias with t(4;11) translocations. *J Proteome Res.* 5 (10), 2743-2753.