CHAPTER 1

INTRODUCTION

1.1) THE ROLE OF *BACILLUS* SP. IN BIOPROCESSING IN ENZYME PRODUCTION

1.1.1 The application of *Bacillus* sp.

Generally speaking, bioprocess optimization aims to improve the outcome of processes that use biological components or living organisms (for instance enzymes or micro-organisms). Bioprocess outcome can be related to quite different criteria, such as maximized product yield in pharmaceutical industry, high product quality in food production, maximum remediation performance in environmental biotechnology or, in a wider context, patient well-being in medicine. Many bioprocesses still have the potential to be improved considerably with respect to these criteria. In order to achieve substantial improvements in bioprocess outcome a systematic approach to bioprocess optimization is required. *Bacillus* sp. continues to be the dominant enzyme-producing microorganisms in applied and industrial microbiology. These organisms are important source of industrial extracellular enzymes, including proteases and amylases, and the biochemistry and physiology of these strains have been investigated in details, Because of the proven capability of these species to produce and secrete gram quantities per litre of these and other enzymes, they have been considered as potential candidate hosts for the production of hetero proteins.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Consequently among Gram-positive bacteria, *B. subtilis* was selected as a priority organism for genomic characterization, the complete genome sequence of *B. subtilis* 168 is now published, comprising 4100 protein-coding gene (1) Of particular relevance to the high capacity of *B. subtilis* and several relatives, especially *B. licheniformis* and *B. amyloliquefaciens*, to secret extracellular proteins was the identification of the several genes encoding proteins of the major secretion pathway: five type I signal peptidase genes and a type II signal peptidase gene, needed for processing lipid-modified precursors, have been found. The sequence confirms the previously characterized intermediary metabolism present in *B. subtilis* consisting of the Embden-Meyerhof-Parnas (EMP) pathway and the triboxylic acid cycle as well as the participation of branched short-chain carboxylic acids in lipid anabolism. The genomic information related to these pathways will provide the tools to facilitate reconstruction of these pathways and to predict patterns of gene expression. Some of the genes of industrial interest those are responsible for the production of carbohydrate-, lipid-, and protein-degrading enzymes, together with some of genes having extracellular protein secretary functions. The analysis of genome of an alkaliphilic *Bacillus* species, *B. halodurans*, assisted by the genomic information from *B. subtilis*, also provided interesting and pertinent information. This strain was first reported as producer of β-galactosidase and xylanase. The rationale for investigating this strain relates to the fact that many commercially useful industrial enzymes are produced from alkaliphilic *Bacillus* strains, and *B. halodurans* has already been well characterized physiologically, biochemical and genetically. While the completion of these genomes is recognized as a major achievement, characterizing the more variable elements of cell make-up, such as the transcriptome (RNA content) and the proteome

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

(reflecting post translational modifications and protein stability), and identifying the nature and role of uncharacterized proteins represent much greater challenges.

1.1.2 The thermostable enzyme from *Bacillus* sp.

Thermostability of enzymes

The roles of enzyme in many processes have been known for a long time. With better knowledge and purification of enzymes the number of applications has increased many folds, and with the availability of thermostable enzymes a number of new possibilities for industrial processes have emerged. Thermostable enzymes, which have been isolated mainly from thermophilic organisms, have found a number of commercial applications because of their overall inherent stability. A number of other applications are in various stages of development. In the food related industry, they have been used in the synthesis of amino acids. In the petroleum, chemical and pulp and paper industries, for example, thermostable enzymes have been used for the elimination of sulphur containing pollutants through the biodegradation of compounds like dibenzothiophene, in the production of 1,3-propanediol from glycerol and in replacing polluting chemical reagents causing toxic products. Currently, a number of publications have extensively discussed developments in this area. Adaptation of extremophiles to hot enzyme form thermophiles and hyperthermophiles, structure and function relationships of thermozymes (heat-tolerant enzymes) (2-4) are among the studied. Thermophile can be alive in very high temperature; these properties imply extremely important industrial and biotechnological implications due to the fact that

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

enzymes from such microorganisms can be employed for use in harsh industrial conditions where their specific catalytic activity is retained. Since thermophiles resist to high temperatures and denaturation, this has the potential to be applied these microorganisms. Microorganisms, like all living things, adapt to the condition in which they have to live and survive. Thermophiles are reported to contain proteins which are thermostable and resist denaturation and proteolysis (5). Specialized proteins known as chaperonins are produced by these organisms, which help, after their denaturation to refold the proteins to their native form and restore their functions (6) The cell membrane of thermophiles is made up of saturated fatty acids. The fatty acid provides a hydrophobic environment for the cell and keeps the cell rigid enough to live at elevated temperatures. The DNA of thermophiles contains a reverse DNA gyrase which produces positive super coils in the DNA (7). This raises the melting point of DNA (The temperature at which the strands of the double helix separate) to at least as high as the organisms maximum temperature for growth. Thermophiles also tolerate high temperature by using increased interactions that non-thermotolerant organisms use, namely, electrostatic, disulphide bridge and hydrophobic interactions. Thermostable enzymes are stable and active at temperatures which are even higher than the optimum temperatures for the growth of microorganisms.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Thermostable proteases

Proteases, which are generally classified into two categories (exopeptidases that cleave off amino acids form the ends of the protein chain and endopeptidases, which cleave peptide bonds within the protein) are becoming major industrial

enzymes, and constitute more than 65% of the world market (8) These enzymes are extensively used in the food, pharmaceutical, leather and textile industries (9-11). The applications will keep increasing in the future as will the need for stable biocatalysts capable of withstanding harsh conditions of operation. Relative ease of the isolation of *Bacillus* from diverse sources has made these organisms the focus of attention in biotechnology (12) So far, however, few thermophilic *Bacillus* sp. that produce proteases have been isolated, the earliest isolate being *Bacillus stearothermophilus* (13) which is stable at 60°C (14), while a different *Bacillus stearothermophilus* sp. produced an alkaline and thermostable protease which is optimally active 85°C (15-16). A species of *Bacillus stearothermophilus* TP26 that has been isolated produces an extracellular protease having an optimum temperature of 75°C (17). Enhancement of protease acitivity excreted form *B. stearothermophilus* had also been possible using economical chemical additives in the proteolysis reactions involved in waste activated sludge. In a chemically defined medium, thermophilic and alkaliphilic *Bacillus* sp. JB-99 was also reported to produce thermostable alkaline proteases (18). Extremely thermostable serine proteases are produced by the hyperthermophilic archaeum *Desulfurococcus* (19) and thermostable metalloproteases are reported from a gram-negative thermophilic bacterium. Major area of focus on the future concerning the production of protease is the optimization of media has a great advantage over complex media in that consistency of processes and production is enhanced through avoiding the variability of complex media in that consistency of processes and production is enhanced through avoiding the variability of complex substrates. Thus, synthetic media provides better control and monitoring, improved product recovery and quality, and simplified purification systems (20). It is for this reason that reports

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

on the production of protease enzymes using synthetic media are available recently (21-22). There is considerable current interest on the exploration of proteases that can catalyze reactions in cold (23). This will allow their use in detergents which can be used in normal tap water without the requirement for increasing the temperature of the water. The search of such enzymes is very much a challenge at this time.

Heat stable lipases

Lipases of microbial origin are the most versatile enzymes and are known to bring about a range of bioconversion reactions which includes hydrolysis, interesterifcaition, esterification, alcoholysis, acidolysis and aminolysis. The esters produced play a relevant role in the food industry as flavor and aroma constituents. Whereas long chain methyl and ethyl esters of carboxylic moieties provide valuable oleo-chemical species that may function as fuel for diesel engines, esters of long chain carboxylic acid and alcohol moieties (waxes) have applications as lubricants and additives in cosmetic formulations (24). Other applications include the removal of the pitch from pulp produced in the paper industry, for the hydrolysis of milk fat in the dairy industry, removal of non-cellulosic impurities for raw cotton before further processing into dyed and finished products, drug formulations in the pharmaceutical industry an in the removal of subcutaneous fat in the leather industry(25). A biodiesel was derived from vegetable oils using immobilized *Candida antarctica* lipase (26). Most of the industrial processes in which lipases are employed function at temperatures exceeding 45°C. the enzymes, thus, are required to exhibit an optimum temperature of around 50°C (27). According to the previous study reported that there

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

are fats exhibiting higher melting points and which are able to inhibit enzymatic reactions at a low temperature (28). Some enzymatic processes for the physical refining of seed oil at about 75°C (29). These reactions, therefore, are enhanced through the utilization of thermo-tolerent lipases.

Lipases are the widespread occurrence thorough the earth's flora and fauna. More abundantly, however, they are found in bacteria, fungi, and yeasts (30). Several *Bacillus* sp. were reported to be the main source of lypolytic enzymes (31). While most of these enzymes are active at a temperature of 60°C and pH of 7.0, lipases from *Bacillus thermoleovorans* and a thermophilic *Rhizopus oryzae* strain can moderately function at extreme pH and temperature values (28). The other problem associated with the production of lipases is required for improvement their secretion at the late stage of growth (32). Proteases secreted in the mean-time either change in the properties of the lipase produced or degraded. Perhaps it is for this reason that only few lipases can be obtained in industrial quantities (30). Hence, the more stable lipase enzyme production in large quantities are required.

1.2) PROTEOMICS

1.2.1 The meanings and scope of proteomics

Proteomics is the study of the proteome, the protein complement of the genome. The terms" proteomics" and "proteome" were coined by Marc Wilkins and colleagues in the early 1990s and mirror the terms "genomics" and "genome", which described the entire collection of genes in organism. These "-omics" terms symbolize a

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

redefinition of how the biology of living systems work. Until the mid-1990s, biochemists, molecular biologists, and cell biologists studied individual genes and proteins or small clusters of related components of specific biochemical pathways. The proteome in any cell thus represents some subset of all possible gene products. However, this does not mean that the proteome is simpler than genome. In fact, the opposite is certainly true. Any protein, though a product of a single gene, may exist in multiple forms that vary within a particular cell or between different cells. Indeed, most proteins exist in several modified forms. These modifications affect protein structure, localization, function, and turnover. However, nowadays, proteomics have several branches which provided the diverse studies of proteomics. For instance, protein separation, it is based on all proteomic technologies rely on the ability to separate a complex mixture so that individual proteins are more easily processed with other techniques. Protein modification is a branch of proteomics which is rely on specialized methods were developed to study such as phosphorylation (phosphoproteomics), glycosylation (glycoproteomics) and redox proteomics. Furthermore, cellular proteomics, a new branch of proteomics, has a goal to map the location of proteins and protein-protein interactions in whole cells during key cell events. Nevertheless, proteomics require several techniques in description of large-scale changes of proteins in any organism. Two dimensional electrophoresis (2DE) is most classical technique in protein separation. Mass spectrometry is also the high-throughput technique for protein identification.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

1.2.2 The tools of proteomics

1.2.2.1 Two dimensional electrophoresis (2DE)

This separation method has become synonymous with proteomics and remains the single best method for resolving highly complex protein mixtures. 2DE is actually a combination of two different types of protein separations. In the first, the proteins are dissolved on the basis of isoelectric point by isoelectric focusing (IEF) (Figure 1.1). In the second focused proteins then are further resolved by electrophoresis on a polyacrylamide gel. Thus 2D-SDS-PAGE resolves proteins in the first dimension by isoelectric point and in the second dimension by molecular weight. Although 2DE is the most effective means of resolving complex protein mixtures it was not widely used for many years after it was first induced in early 1970s. 2D-SDS-PAGE systems that use immobilized pH gradient (IPG) strips and relatively foolproof hardware to facilitate the transfer of proteins from the IPG strip into SDS-PAGE slab the transfer of proteins from IPG strip into the SDS-PAGE slab gel. The IPG strip is based on the use of immobilized pH gradients. One can now purchase IPG strips from major suppliers that afford reproducible separations over a variety of wide and narrow pH ranges. The use of narrow pH ranges facilitates the separation of proteins with highly similar isoelectric points. The strip is hydrated with a buffer and the proteins are slowly loaded into the strip under voltage. Then the voltage is increased to achieve focusing.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

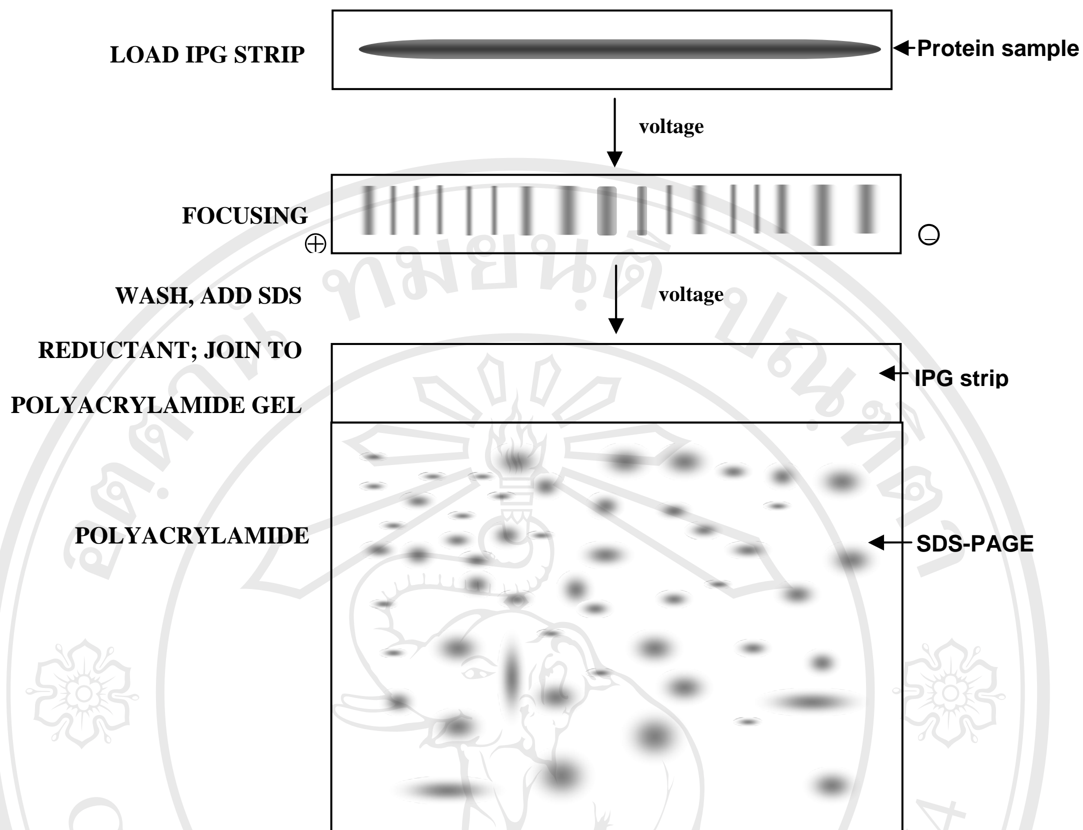


Figure 1.1 Schematic representation of 2D-SDS-PAGE

Commercially available systems provide temperature control as well as highly accurate voltage or current control to facilitate reproducible separations.

After the focusing step, the strip is treated with a buffer that contains a thiol reductant and SDS and then is joined to the SDS-PAGE slab gel. In the respect, the IPG strip containing the focused proteins acts as a "stacking" gel in 1D-SDS-PAGE. The proteins then are resolved on the SDS-PAGE slab gel in the same manner as for 1D-SDS-PAGE.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Proteins separated by 2D gels are visualized by conventional staining techniques, including Silver, Coomassie, and Amido black stains. Silver-staining and newer fluorescent dyes are the most sensitive. Although there are many different protocols for all of these staining of the proteins. For example, Silver-staining with formalin fixation of the proteins tends to fix proteins in the gel, preventing both their digestion and the recovery of any peptides formed. Similar problems result form prolonged exposure of gels to acetic acid. Thus, it is important to use staining protocols that are compatible with subsequent digestion and elution steps.

1.2.2.2 Comparative proteomics with 2DE gel image analysis

Perhaps the most widely used approach to comparative proteomics is to subject two samples to 2D-SDS-PAGE and compare the spot patterns. Two-dimensional SDS-PAGE is particularly well-suited to comparative proteome analysis because it effectively resolves many proteins. With recent improvements in 2D gel technology, the task of running reproducible 2D gels has been made easier. Even before the introduction of MS-based protein identification, this approach provided a useful means of comparing proteomes. However, identification of the proteins was cumbersome and difficult. Application of peptide mass fingerprinting and LC-MS/MS analyses now makes it possible to identify essentially any protein one can detect by staining the gel. Thus, the critical task in comparative proteomics with 2D gels is identifying the features that differ between gels (Figure 1.2).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

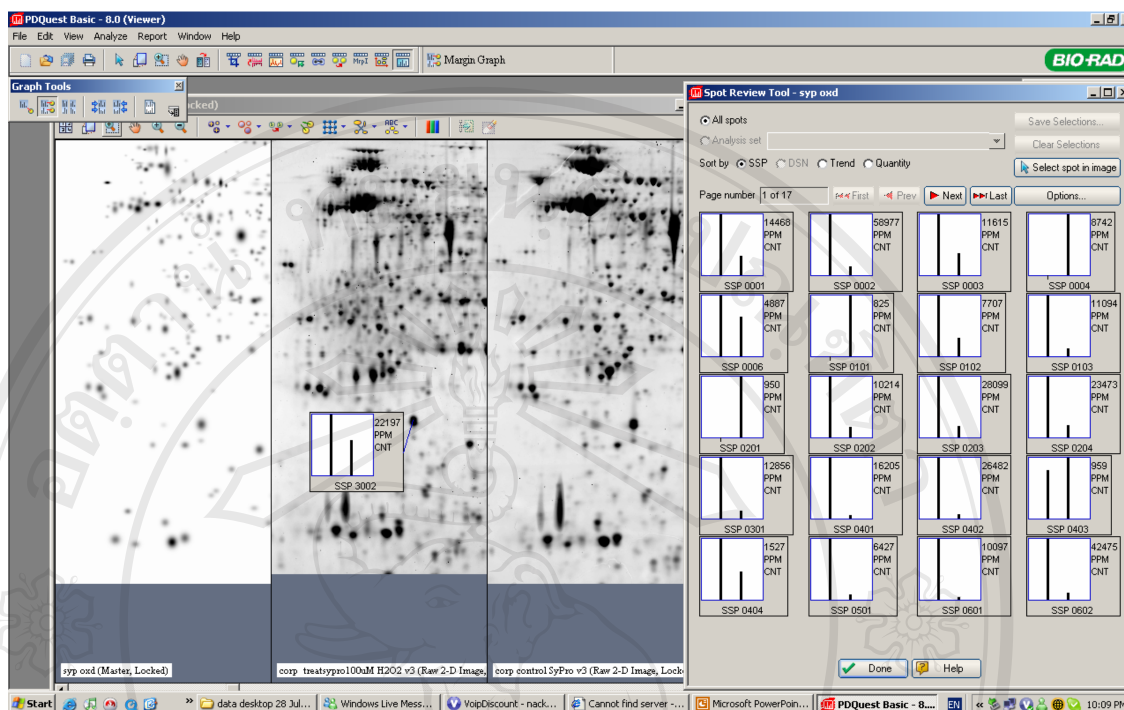


Figure 1.2 Quantitation and comparison of spot intensities on multiple 2DE gels with PDQUEST 2-D analysis software.

A great deal of work has been done to develop software tools to analyze the patterns of protein spots on 2D gels. In addition, extensive databases to archive this information have been developed. Among the most widely used programs for 2D gel image analysis is PDQUEST 2-D analysis software. This software works with images of stained 2D gels. These images can be acquired by the use of CCD camera. The program does several things. First, the gel is evaluated for “features”, which simply refer to any significant deviation from the background. The features correspond to the protein spots on the gels. The volume was derived from the integration of OD over the spot area. These characteristics comprise the basis for comparing features within a gel

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

and between multiple gels. Of course, for protein-expression profiling, the 2D gels from two different samples are required to compare for differences in the occurrence of intensity features. The problem with this is that it is very hard to run multiple 2Dgels with exact reproducibility. There usually are slight variations in the location of spots for specific proteins. This makes the user identify "landmarks" which are proteins that are "paired" by the software to create a series of pairs by which the gels can be aligned or "matched". The matching process involves to each other in 2D space. In other words, the gel images are lined up entail some transformations or spatial "wrapping" of images to compensate for local geometric distortions in the gel. Once the gels are matched, then comparison of the features may be done. These comparisons examine the OD volume differences between features on the gels and provide a graphical output that assigns numbers to the observed differences. The software also enables statistical analyses of these data to facilitate interpretation of significant differences. It is this operation that allows the user to identify those features or spots that differ between two or more samples. Gels may be visually "stacked" to enable comparison of images. Alternatively, virtual gels can be synthesized from the images collected from multiple gels to provide a master archive of composite proteomes in different states of an organism.

The use of 2D gels is a powerful approach to protein profiling and it is unique in providing a visual-image basis for proteome comparisons. However, there is one major drawback to this approach: staining of 2D gels only detects the more abundant proteins in a sample. There is approximately a million-fold range of protein expression in cells whereas gel staining is limited by about a hundred-fold dynamic range. It is possible to enhance detection of low-abundance proteins by loading more

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

protein for analysis, but abundant proteins eventually overwhelm many of the features on the gel. A related problem is that many proteins exist in multiply modified forms, which may display different isoelectric points and are thus separated on 2D gels. For less abundant proteins, spreading out into multiple spots can lower detectability by staining. Finally, identification of very weakly stained proteins by in-gel digestion is hampered by poor recovery of peptides form the digestion and the gel. Although staining and visualization methods are continuing to evolve and improve, this problem may ultimately limit 2D gels to analysis of relatively abundant proteins. This is adequate for many circumstances, however, and 2D gel-based proteome profiling will continue to be valuable, widely used technique.

1.2.2.3 Protein identification

Peptide mass fingerprinting is a protein identification technique in which MS is used to measure the masses of proteolytic peptide fragments. The protein then is identified by matching the measured peptide masses to corresponding peptide masses from protein or nucleotide sequence databases. Peptide mass fingerprinting works well for analytical proteomics because it combines a conceptually simple approach with robust, high-throughput instrument (typically MALDI-TOF MS). As with other MS-based analytical proteomics techniques, the quality of the protein identifications made depend on the quality of both the MS data, the accuracy of the databases, and the power of the search algorithms and software used. Modern MS instruments are capable of measuring the molecular weight of intact proteins with a fairly high degree

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

of accuracy. Thus, the interesting protein spots will be excised and the protein will be digested with trypsin or other proteolytic enzymes.

1.2.2.3.1 In-gel Digestion

A commonly used appoarch to digestion of proteins separated by 1D- or 2D-SDS PAGE is referred to as "in-gel" digestion (Figure 1.3). The band or spot of interests is cut from the gel, destained, and then treated with protease (most commonly trypsin). The enzyme penetrates the gel matrix and digests the proteins to peptides, which then are eluted from the gel by washing. This technique is an indispensable element to 2D-SDS-PAGE proteomics strategies. Although trypsin is the most commonly used enzyme, the general approach is applicable to other proteases, including Glu C and chymotrypsin. The efficiency of both digestion and recovery of peptide from the gels is highly variable. A key determinant of successful in-gel digestions is the gel-staining technique used. Staining protocols that employ aldehyde fixative or prolonged exposure to acids (e.g. acetic acid) tend to fix proteins in gels, thus making the proteins difficult to digest and the peptides difficult to elute.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

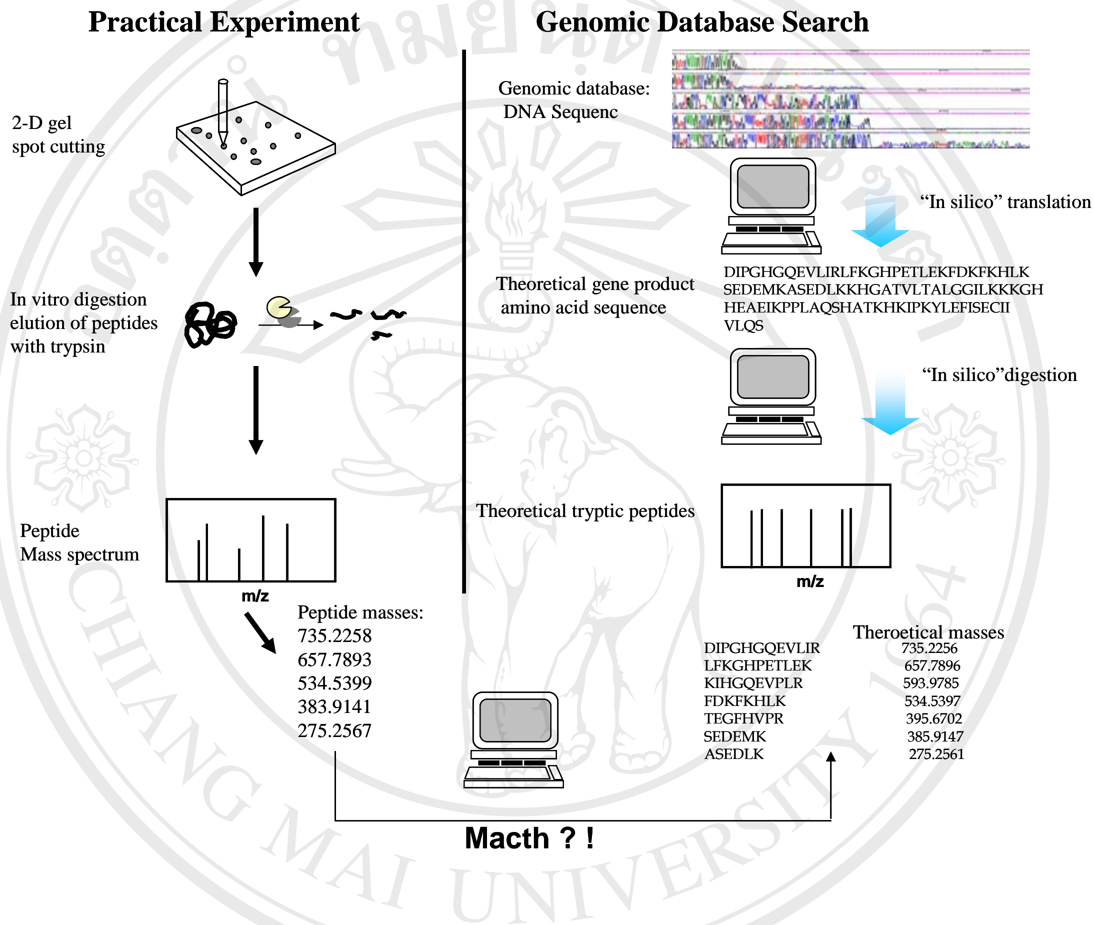


Figure 1.3 Protein identification with peptide mass fingerprinting The peptide masses of the digested protein are matched with a list of theoretical masses of peptides which are mathematically derived from the open reading frames of the genome database of a certain organism.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

With highly cross-linked gels, the penetration of protease enzymes into the gel matrix may be retarded. Finally, residual components of the SDS-PAGE technique (SDS or residual unpolymerized acrylamide) may be inhibitory to protease activities.

1.2.2.3.2 Matrix Assisted Laser Desorption Ionisation (MALDI) principle

MALDI ions are created by missing the analyte with a small organic molecule which absorbs light at the wavelength of the laser the matrix. The analyte becomes incorporated into the crystal lactice of the matrix and is then irradiated with a laser. The laser causes the desorption and ionization of the matrix and analyte, either by protonation or cationation (positively charged ions) or by deprotonation (negatively charged ions). The ions are then accelerated into the MS analyzer. The typical wavelength of the UV lasers utilized is 337 nm. The precise mechanisms of desorption and ionization is still unclear. Generally irradiation of the matrix-analyte solid causes sublimation of the matrix-analyte solid and ejection of a plume of matrix and analyte ions into the gas phase. MALDI ions are generated under high vacuum ($5x10^{-6}$). Packets of ions are generated with each laser pulse and each packet is pulsed into the analyzer. As MALDI is a pulsed ionization technique, it is ideally coupled with a TOF analyzer. A wide range of matrices for bio-mass spectrometry applications have been adopted for use with UV lasers. The three widely used matrices for peptides and proteins are α-cyano, 4-hydroxy cinnamic acid for peptide and it is still widely used today affording high sensitivity and negligible matrix adduction. Many different procedures for the sample preparation for α-cyano exist.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

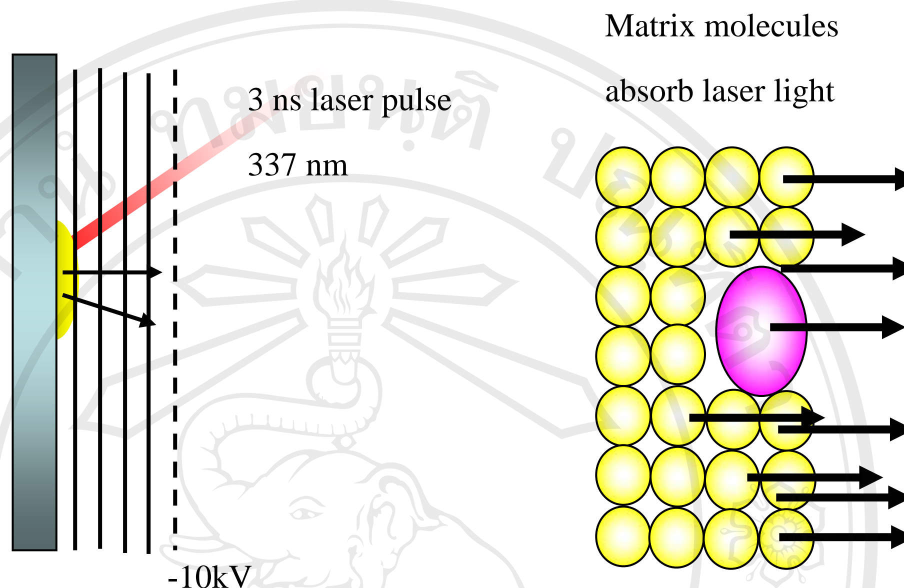


Figure 1.4 Schematic of MALDI process

1.2.2.3.3 Tandem mass spectrometry (MS/MS)

In tandem mass spectrometry or MS/MS, the analyte ion of interest is measured, specifically selected and fragmented in the mass spectrometry by collision induced dissociation (CID), generating structural information. The technique is performed with instruments capable of selecting ions of particular *m/z* value and subjecting the selected ions to fragmentation within the mass spectrometer.

Most MS/MS instruments consist of two mass analyzers arranged in tandem, but separated by a collision cell. In a MS/MS instrument, sample ions of a specified m/z value can be selected by the first analyzer and then directed into the collision cell

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

where they collide with neutral gas molecules. The use of a collision cell means that ion fragmentation is induced deliberately and in a specific region in the instrument.

Mass analyzer

The function of mass analyzer is to separate ions according to their mass-to-charge (m/z) ratio

Time of flight analyzer (TOF)

Time-of-flight analyzers are one of the simplest MS analyzers in use today. Ions produced in the ion source are accelerated by high voltage into the TOF analyzer acquiring an initial velocity that is dependent on their masses. Mass measurement is recorded by the time of flight of an ion in the ToF flight tube, the time of flight of an ion is proportional to the square root of its mass/charge ratio, given a constant accelerating voltage.

$$\text{Time of flight} = k\sqrt{m/s}$$

The time of flight of an ion is typically measured as the time it takes to traverse from the flight tube to the detector tube when the ion leaves the sources.

Quadrupole analyzer

The most commonly used analyzer is the quadrupole mass filter. In this device, a voltage made up of a DC component U and an RF component. V cos ωt is applied between adjacent rods of the quadrupole assembly, whereas opposite rods are connected electrically. With a correct choice of voltage, only ions of a given *m/z* value

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

can traverse the analyzer to the detector, whereas ions having other *m/z* values collide with the rods and are lost. By scanning the DC and RF voltages, while keeping their ratio constant, ions with different *m/z* ratios will pass successively through the analyzer. In this way, the whole *m/z* range may be scanned and a complete mass spectrum recorded.

1.2.2.3.4 Peptide Mass Fingerprinting

The technique originally described in 1993 comprises protein digestion, MALDI-ToF analysis and sequence database search algorithms (Figure 1.5).

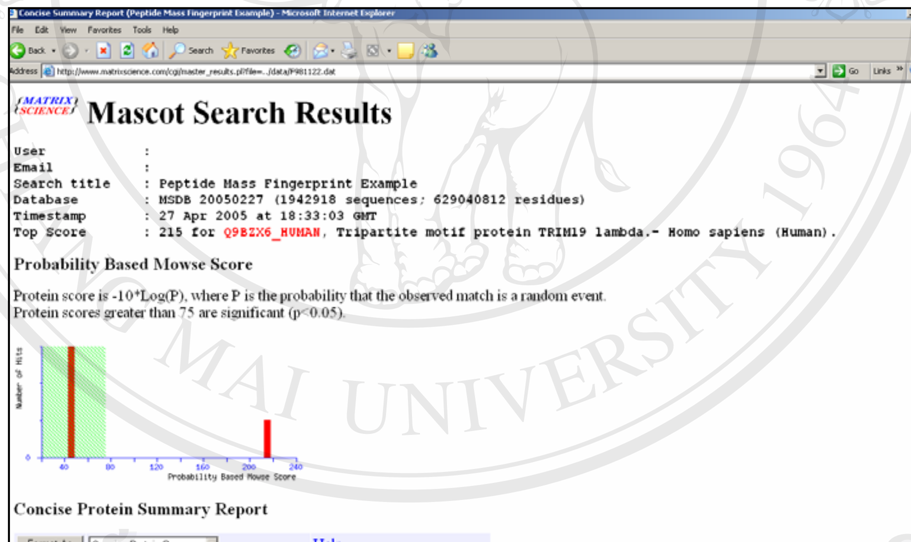


Figure 1.5 The identified protein result from MASCOT program from www.matrixscience.com

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Simply, every protein databases or genomic database is theoretically digested with the cleavage reagent used in the digestion reaction, generating of hundreds to thousands of theoretical peptides. The experimental peptide masses derived from the MS spectrum, the peptide mass fingerprint (PMF) is then compared to the theoretical peptide masses and score is calculated and assigned. The score reflects the match between the theoretically and experimentally determined masses, the identified as the most probable is the one that gives the best match between the experimental and the theoretical peptides. The number of peptides observed in the PMF and the accuracy to which they are measured determines the confidence of the protein identification. The incorporation of reflectron technology and delayed extraction into MALDI-TOF instrumentation has enhanced the performance of peptide mass fingerprinting considerably. Several programs are available to perform this type of search, varying in execution of the task (including MASCOT at www.matrixscience.com profound at www.prowl.com MS-FIT at www.prospector.ucsf.edu) Accuracy, reliability and speed will determine the program of choice. Regradless which program is used, four user variables are important for a PMF search 1) peptide mass list 2) Specification of the cleavage agent 3) Error tolerance. The accuracy of mass measurement is determined by the calibration, the higher the mass accuracy the greater the specificity 4) Knowledge of peptide modification e.g. methionine oxidation.

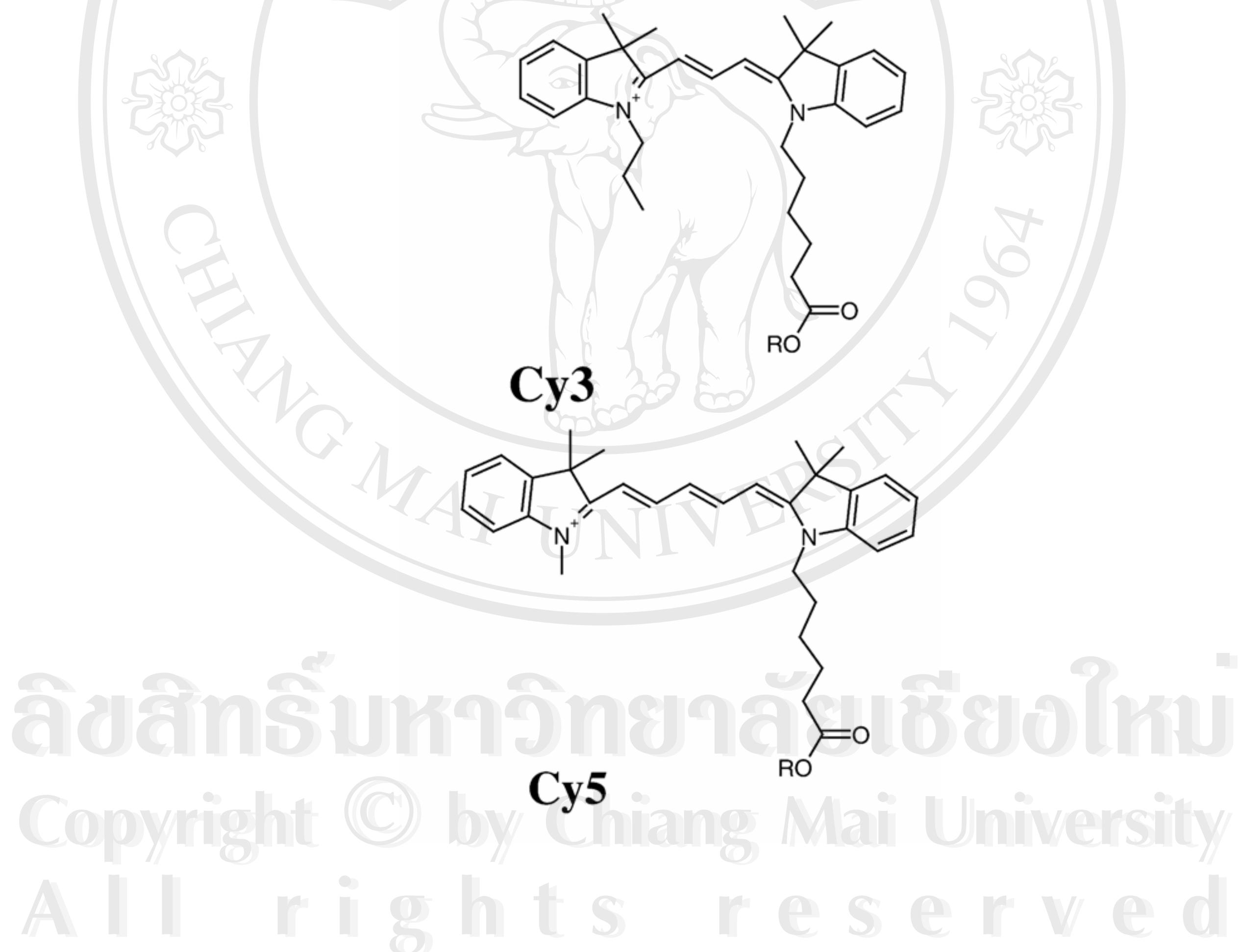
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

1.2.3 Two dimensional different in-gel electrophoresis (2D-DIGE)

A mainstay of conventional proteomics is high-resolution 2DE followed by protein identification using mass spectrometry. The state-of-art 2D-gel system can be loaded with a few milligrams of protein and separates thousands of protein spots. While the technique has been widely used and successfully applied in a variety of biological systems, several technical limitations exist. Due to 2DE gel usually can not be fully duplicated, which makes it difficult to find the proteins changed between gels and to quantify the changes in protein expression. While a comparison of protein expression profiles from regular 2DE gel can be carried out with the assistance of various software programs, it typically requires some computerized justification of 2DE gel images so that two images can be superimposed and compared. These difficulties limit the speed and accuracy of quantitation of protein spots in 2DE-gel electrophoresis. The Differential In-Gel Electrophoresis (DIGE) technique recently introduced by Amersham Pharmacia Biotech is aimed at improving reproducibility. To analyze the sample in DIGE, two pools of protein extracts are labeled covalently with fluorescent cyanide dyes, Cy3 and Cy5, respectively (Figure1.6). In general, this CyDye is based on differential labeling with N-hydroxy-succinimide ester-modified cyanine which reacts with ε-amino group of lysine. Cy3 is excited at 540 nm and has emission maximum at 590 nm while Cy5 is excited at 620 nm, and emits at 680 nm. The molecular masses of the dyes are similar to each other (434 and 464 Da). These labeled proteins are mixed and separated in the same 2DE-gel (Figure 1.7). The 2DE gel patterns can be rapidly images by the fluorescence excitation of either Cy3 or Cy5 dyes. The amount of the dye is controlled in such a way that on average one protein

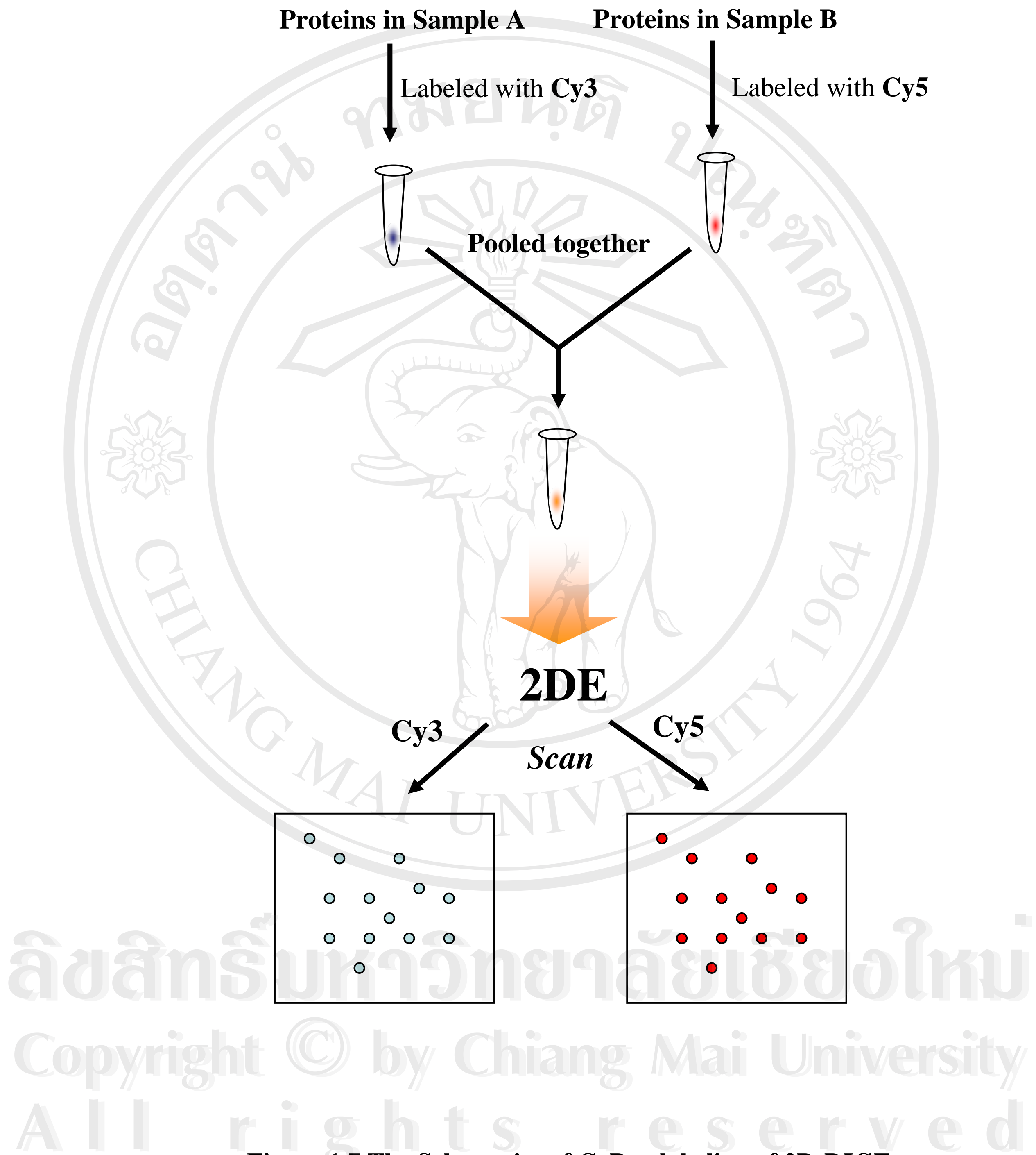
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

molecule is labeled not more than once and minimum number of the molecules of each protein labeled. A comparison of the resulting image allows quantitation of each protein spot. Since two pools of the proteins are separated in the same gel. Those proteins existing in both pools will migrate to the same locations in the 2DE gel, minimizing the reproducibility problem. Quantitation of the protein profile can be rapidly and accurately achieved based on the fluorescence intensity.



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Figure 1.6 Chemical formulate of Cy3 and Cy5 N-hydroxy-succinimide ester.

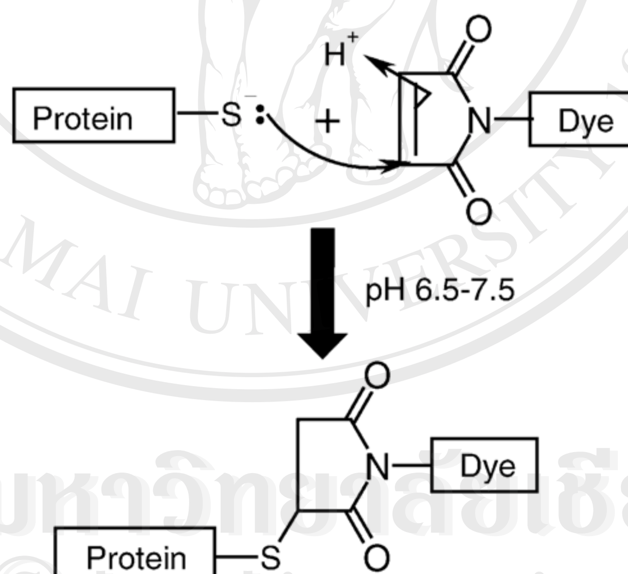


ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Figure 1.7 The Schematics of CyDye labeling of 2D-DIGE

However, this study applied this DIGE strategy for detection of thiol oxidized proteins by using another protocol for differential Cy3/Cy5 labeling, based on the

reaction of a similar set dyes not any longer Lysine, but on Cysteine residues. A labeling protocol aimed at Cysteine residues would automatically extinguish any further reactivity. As shown in figure 1.8, which gives the type of reaction of such compounds, it must be stated that the reaching tail of these two fluorophores is also quite appropriate in 2DE analysis, since it is not an iodinated tail, which would automatically be destroyed by the thiourea in the solubilizing medium. The reacting end is indeed a Maleimide residue, permitting an addition of the –SH group to the double bond of the Maleimide moiety, thus forming a thioether link. It has suggested that α-β unsaturated compounds should be preferred as alkylating agents for-SH groups in lieu of iodoacetamide.



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Figure 1.8 Reaction scheme of Maleimide cyanide dyes with the –SH group of proteins

1.3) POST-TRANSLATIONAL MODIFICATION

Post-translational modification is the chemical modification of a protein after its translation. The posttranslational modification (PTM) of amino acids extends the range of functions of the protein by attaching to it other biochemical functional groups such as acetate, phosphate, various lipids and carbohydrates, by changing the chemical nature of an amino acid (e.g. citrullination) or by making structural changes, like the formation of disulfide bridges. Also, enzymes may remove amino acids from the amino end of the protein, or cut the peptide chain in the middle. For instance, the peptide hormone insulin is cut twice after disulfide bonds are formed, and a propeptide is removed from the middle of the chain; the resulting protein consists of two polypeptide chains connected by disulfide bonds. Other modifications, like phosphorylation, are part of common mechanisms for controlling the behavior of a protein, for instance activating or inactivating an enzyme.

1.3.1 Thiol modifcation

As the premier biological electron acceptor, molecular oxygen (O_2^-) serves as vital role in fundamental cellular functions, including the process of aerobic respiration. Nevertheless, with the beneficial properties of O_2 comes the inadvertent formation of reactive oxygen species, including superoxide (O_2^-), hydrogen peroxide (H_2O_2), and hydroxyl radical ($^{\cdot}OH$). These differ from O_2 in having one, two, and three additional electrons, respectively (Figure 1.9).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

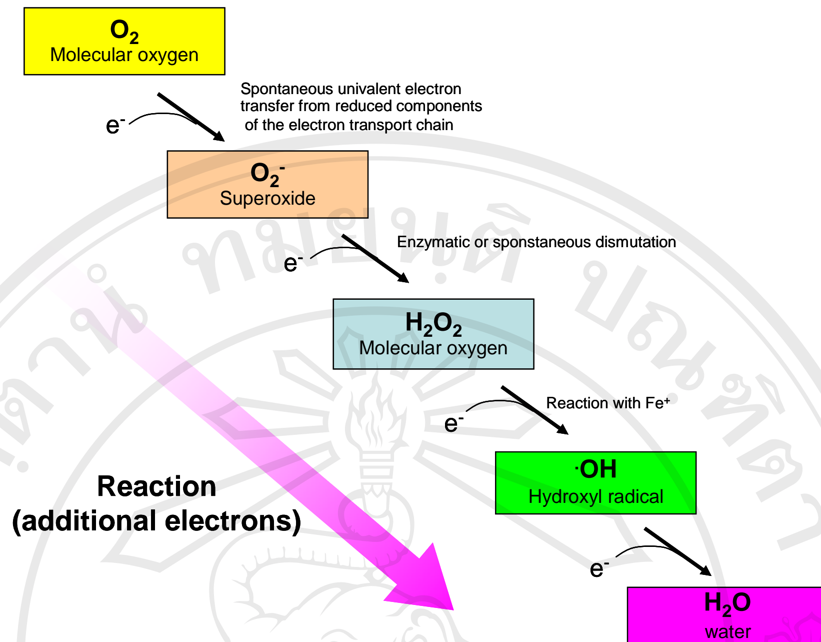


Figure 1.9 Formation of Reactive Oxygen Species

Cells also encounter elevated levels of these reactive oxygen species when they are released by animals, plants and insects as a defense against detrimental organism such as microbial pathogens. Reactive oxygen species of lipids, which leads in turn to perturbations in membrane structure and function. The accumulation of this oxidative damage underlies the formation of man disease states in humans. It is postulated that tissue injury by these reactive oxygen species accumulates over a long period of time and plays roles in the aging process and the development of heart disease, diabetes, chronic inflammatory diseases, cancer, and several neurodegenerative disease. Many organisms have evolved strategies and repair damage, which have enabled them to prosper from the tremendous oxidizing potential of O_2 without succumbing to

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

oxidative damage. Bacteria, yeast and mammalian cells all induce the synthesis of global regulatory responses to survive oxidative stress. Furthermore, the consequences of oxidative stress and the corresponding defense responses have been extensively studied in *E. coli*. The stress responses are often provoked by the external addition of chemical oxidants that specifically elevate the levels of reactive oxygen species with cells, or by the use of mutant strains that disrupt the normal "homeostatic mechanisms" for removing reactive oxygen species the damage they do. The major target of O_2^- damage identified in bacteria is a class of dehydratase enzymes that utilize [4Fe-4S] clusters to bind their substrate (33).

Since some of these enzymes function in the citric acid cycle (also called the Kerbs cycle) and in amino acid biosynthesis, high levels of O_2^- lead to a requirement for certain amino acids in growth media (34). H_2O_2 is well known for its role in oxidizing thiol (SH) groups of cysteinyl amino acid residues in proteins. Elevated levels of H_2O_2 also are associated with the oxidation of other amino acids, leading to the formation of methionine sulfoxide and a variety of carbonyls. Lastly, because of its extreme reactivity OH targets all of the major macromolecules of cells, RNA, DNA, protein and lipids. The extent to which membrane lipids are targets appears to depend on the presence of polyunsaturated fatty acids in lipids, which are not as prevalent in bacteria as they are in mammals. Many enzymes that protect against oxidative damage have been identified in *E. coil* (33, 35). Three superoxide dismutases, each of which contains a different metal center and show different expression patterns and subcellular localization, catalyze the dismutation of O_2 to H_2O_2. While the superoxide dismutases eliminate O_2, they also are a source of endogenously produced H_2O_2 in *E. coli*. The major enzymes involved in reducing

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

H_2O_2 to H_2O and O_2 in *E. coli* are catalase and alkyl hydroperoxide reductase. There is no enzymatic mechanism for decreasing levels of OH, produced from H_2O_2. Thus, levels of OH will be directly proportional to levels H_2O_2, and accordingly, catalase and alkyl hydroxide reductase activities are critical to oxidative stress survival. Another component to the oxidative stress response is the reduction of oxidized thiols that arises through one of the mechanisms described below, The tripeptide glutathione and the thiol reductants glutaredoxin and thioredoxin are key to the restoration of thiols to their reduced state (-SH) (36). *E. coli* contains three glutaredoxins that utilize the reducing power of glutathione to catalyze the reduction of disulfide bonds (-S-S-) in the presence of NADPH and glutathione reductase. There are two thioredoxins in *E. coli* that also function to reduce disulfide bonds. Reduced thioredoxin is regenerated by thioredoxin reductase and NADPH. The fact that NADPH is required to maintain the reduced state of glutathione and thioredoxin indicates that the response to oxidative stress is coupled to the physiological status of core pathways that generate NADPH.

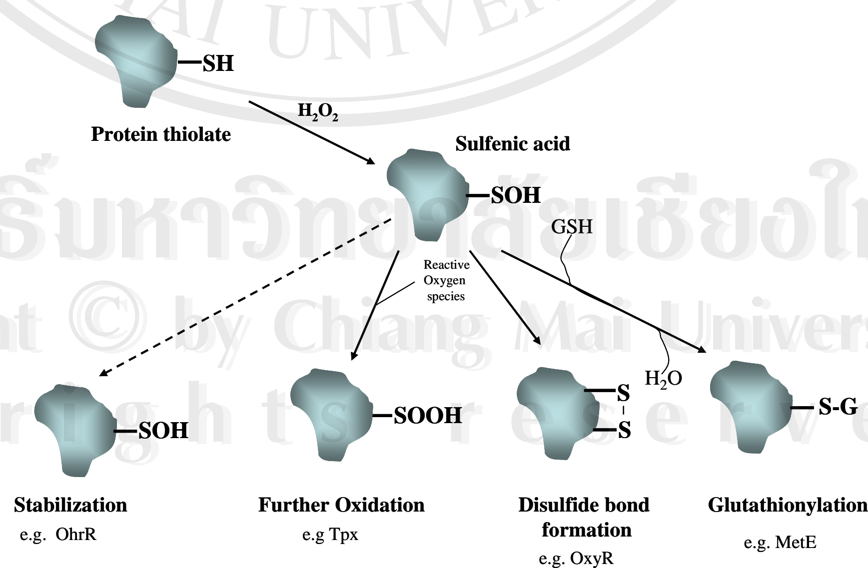


Figure 1.10 Thiol modifications of proteins.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Formation of sulfenic acid from the reactive ion of H_2O_2 with protein thiolates leads to different protein modifications, depending on the protein. In proteins without a second sulfyhydryl, the sulfenic acid (-SOH) may be stabilized (e.g. OhrR) or may react with reactive oxygen species to generate the further oxidatized sulfinic (-SO_2H) (e.g. thiolperoxidase, Tpx) and sulfonic acid (-SO_3H) derivatives. Alternatively, if a second cysteinyl residue is in proximity within the same polypeptide (e.g. OxyR) or an associated protein (e.g. Yap1 and Orp1), a disulfide bond can form between the two sulfur atoms (-S-S-). Lastly, the sulfenated cysteinyl residue can react with glutathione (GSH), leading to a mixed disulfide (e.g MetE).

1.3.2 Regulatory Roles of Thiol Modifications

As mentioned above, proteins-particularly, the thiols of cysteines are the major targets of H_2O_2 can lead to the formation of different modifications such as sulfenic acid (-SOH), sulfinic acid (-SO_2) and sulfonic acid (-SO_3H) as well as disulfide bond formation (-S-S-) and glutathione conjugation(-S-GSH) (figure 1.10) (36-37). These modifications often alter the structure and function of protein. Recent progress in this field points to a common chemistry in the reaction of H_2O_2 with thiolates through the initial formation of sulfenic acid. In the case of proteins that have a nearby cysteinyl residue, a disulfide bond forms between the two sulfur atoms The sulfenated cysteinyl residue also can react with a cysteinyl residue on another protein or with glutathione,leading to a mixed disulfide. If no cysteinyl residue is nearby, the sulfenated cysteine can be further oxidized to sulfinic or sulfonic acid or it can remain in the sulfenic acid state. All but the sulfinic and sulfonic acid modifications are

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

readily reversible by reduction,using proteins such as thioredoxin or glutaredoxin, though sulfinic acid reductase activities have recently been identified in yeast and mammalian cell(denoted sulfiredoxin and sestrin, respectively) (39-40).

Given the reversible nature of most forms of thiol oxidation, it has been suggested that thiol modifications can play roles in signal transduction that are similar to protein phosphorylation/dephosphorylation. In support of this model, there are several examples of proteins whose activities are modulated by thiol oxidation and reduction.

The first of these examples is the OxyR transcription factor, which up-regulates peroxide defenses in *E. coli* and a variety of other bacteria. OxyR contains two critical cysteines that are oxidized to form an intramolecular disulfide bond when cell encounter peroxide stress (41-42). Disulfide bond formation is associated with a conformational change that alters OxyR binding to DNA and allows the protein to activate the transcription of genes encoding enzymes, such as catalase and the alkyl hydrogenperoxide reductase, that destroy H_2O_2. Once the H_2O_2 concentration is decreased, OxyR is reduced and the system is reset. The unusually reactive cysteine in OxyR that is oxidized by H_2O_2 to form the sulfenic acid intermediate can clearly be nitrosylated and glutathionylated in vitro (43, 44). but the in vivo relevance of these other modifications is questionable (45).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Two other examples of redox regulated proteins are the *E.coli* chaperone protein Hsp33 (46) and the *Streptomyces coelicolor* anti-sigma factor, RsrA (47-49). For these proteins, the cysteine residues, which form intramolecular disulfide bonds, are in a reduced state when coordinated to a zinc (Zn^{2+}), and zinc is released upon

oxidation of the thiols. For both proteins, oxidation and zinc release are associated with an opening of the protein structure. For Hsp33, this structural change allows for dimerization and activates its chaperone activity (50). RsrA, on the other hand, dissociates from a promoter specificity factor of RNA polymerase (an extracytoplasmic-function-type alternative sigma factor) allowing the transcription of genes that permit recovery from the stress (47). Among the target gene products is a thioredoxin, which reduces the disulfide bonds that form within oxidized RsrA. Presumably, reduction of the disulfide restores the binding of zinc and its inhibitory association with the sigma factor. Thus, the RsrA regulatory circuit provides another example, comparable to OxyR, in which the modification of a regulatory protein thiol group can be linked to a change in the transcriptional output of genes that remediate stress.

The peroxide-sensing repressor OhrR from *Xanthomonas campestris* pv. *phaseoli* (51) and *Bacillus subtilus* (58) can be inactivated by H_2O_2 or by organic peroxides (ROOH) formed by the oxidation of a variety of organic molecules in the cell or in the environment.The *B.subtilis* OhrR transcription regulator contains only a single cysteine that forms a relatively stable sulfenic acid upon its reaction with H_2O_2 or organic peroxides (52). Oxidation of the single cysteine leads to the dissociation of OhrR from its DNA binding site and the depression of the gene encoding an organic hydroperoxdase that eliminates the initial oxidizing insult.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Hondorp and Matthews (53) provide an example of a thiol modification that protects an enzyme activity during oxidative stress. Their data suggest that when cells encounter oxidative stress, a key cysteinyl residue near the active site of methionine

synthase (MetE) is glutathionylated. This modification blocks access of the substrate and prevents further synthesis of methionine. This finding is significant in that it presents a mechanism to reversibly preserve the function of a protein during oxidative challenge. By glutathionylating a single cysteinyl residue, the protein is protected from further oxidation of that cysteinyl residue to the irreversible sulfinic and sulfonic acid forms. Once the stress is removed, the mixed the stress is removed, the mixed disulfide bond will be readily reduced, and access to the substrate restored.

1.3.3 Prevalence of Regulatory Thiol Modifications

As illustrated by the examples above an array of chemical modifications obtained by oxidizing cysteinyl residues has been exploited in combating oxidative stress. Yet it is important to note that not all cysteinyl residues of proteins are readily oxidized by oxidants such as H_2O_2 (38) The *p*Ka of the thiolates are more reactive than their protonated counterparts. In addition, the contribution of protein environment to the stability of the oxidized products is also known to be a factor, but is not well understood. Given that many of the thiol modifications do not appear to be in equilibrium with the redox state of the cell, the features of the protein that determine the rate at which the modifications are formed are also important parameters. The added complexity of the cysteine targets that compose part of a Zn binding site found for Hsp33 and RsrA raised questions about the function of the zinc. Perhaps Zn binding provides some additional control over the reactivity of the cysteine thiols, or perhaps the loss of the Zn facilitates conformational changes. Recently, the oxidative, stress-induced thioredoxin2 from *E. coli* has also been shown to contain a H_2O_2-labile

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Zn does not changes its reductase activity (53). Thus, the way this oxidatively labile Zn site affects thioredoxin function has yet to be established. The extent of thiol oxidation within the cell remains another open question, The variety of modifications that arise from treatment with H_2O_2 and the experimental challenges associated with their detection has made it difficult to catalog all the proteins that are modified and all the types of modification exist The importance of monitoring transient changes in cysteines are highlighted by the recent finding that oxidation of the Yap1 activator of antioxidant genes in the yeast *Saccharomyces cevevisiae* requires a peroxidase denoted Gpx3 or Orp1 (54). In this case, H_2O_2 reacts with a cysteine in Orp1, forming an unstable sulfenic acid intermediate that then reacts with a cysteinyl residue of Yap. The disulfide undergoes an exchange with a second cysteine within Yap1 to form Yap1 in a confirmation that masks the nuclear export signal (55). Thus, methods that allow the appearance of thiol modifications in cells to be monitoredkinetically will greatly enhance our understanding of how cysteine residues become oxidized. The examples mentioned here illustrate the versatile potential of thiol modifications. Given the reversibility of thiol oxidations and the wide range of structural constrains that can be imposed by the formation of a by the formation of a sulfenic or sulfinic acid or a disulfide bond.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

1.4) THE BACKGROUND OF *BACILLUS STEAROTHERMOPHILUS* TLS33

The thermophilic bacteria are traditional producer of thermostable enzymes. Among the thermophilic bacteria, *Geobacillus stearothermophilus* (formerly, *Bacillus stearothermophilus*) was used as thermostable enzyme producer. The enzymes were discovered form this species, for instance, lipase and protease. In the previous studies, Sinchaikul S. and Sookkheo B. have studies in the characteristics of these enzymes. It was suggested that the lipase from *B. stearothermophilus* P1 is a novel enzyme found in an organism growing in a hot spring in Chiang Mai, Thailand. Sinchaikul S. et al have cloned the thermostable lipase into *E. coli* and purified for studying its characteristics such as molecular weight, stability, inhibitor and substrate specificity, it has been found that this thermostable lipase has molecular weight 43,209 Da by analyzing with MS LCQ (55). The optimum temperature of this lipase is approximately 55-60°C. This lipase hydrolyzes synthetic substrates with acyl group chain length of between C8 and C12, with optimal activity with C10 (*p*-NP-caprate). The lipase activity in long chain of substrates between 70 and 100% of optimal for C8 or C10 groups and 30 and 50% for C12 to C18, whereas, with short-chain substrates (C2-C6), lipase activity was less than 30%. In addition, the lipase hydrolyzed triacylglycerol with acyl-group chain lengths of between C8 and C12, with optimal activity with C8 (tricaprylin). In addition, the lipase hydrolyzed trilinolenin more than trillinolein and triolein. The characterization of the purified lipase was very interesting, and showed that it was active over a wide range pH values from 7-10. Most lipases have a catalytic triad consisting of Ser-His-Asp/Glu (57) similar to that in serine proteases. The catalytic serine is embedded in a signature pentapeptide sequence. Gly-X-Ser-X-Gly, located at the C-terminal of a secretion of parallel

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

strands of β-sheet. Serine, one part of the catalytic site, is embedded in a tight bend between an α-helix and a β-stand. Sookkheo B. *et al* have purified and characterized the thermostable proteases from *Geobacillus stearothermophilus* TLS33. According to the industrial application of thermostable lipase such as detergent, food including pharmaceutical industries, this thermophilic bacterium was attempted to study its protease. They found that there were three proteases and assigned name as size, S, B and N.

1.5) The previous study of *Bacillus subtilis* proteome

With the publication of the first genome sequence of a living organism in 1995 (58) a new era in biology was opened. This era of functional genomics provides for the first time a global and comprehensive view of life in general towards the elucidation of the mystery of life. During the past 10 years, *Bacillus subtilis* has been learned a lot from these genomics data. Comparative genomics combined with bioinformatics, for instance, provides the key to a new understanding of the evaluation of bacteria. The genomic sequence, however, presents only the blue-print of life, not life itself. Now functional genomics is required to bring this static genome information to dynamics of life. The proteome is highly dynamic and flexible and thus reflecting environmental stimuli imposed to the cell. A number of the reports focused on *Bacillus subtilis*, the model organism of Gram positive bacteria. Interest in *Bacillus* sporulation as a value model system for analyzing cell differentiation at a molecular level as well as its industrial potential for the production of many extracellular enzymes have both contributed to its attractiveness and its development into a model

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

organism. Therefore, an extensive knowledges of the genetics, biochemistry, and physiology of *B. subtilis*. The combination of the sophisticated molecular tools, an extensive database and the toolbox of functional genomics makes functional genomics approaches in *Bacilli* particularly rewarding. Sequencing of the *B. subtilis* genome revealed about 4100 genes including 1700 genes with still unknown function (59) thus indicating that many chapters of the "Bible of *bacillus*", one of the most intensively studied organisms at all, are still empty (60). The elucidation of the function of this surprisingly high number of unknown genes is a big challenge for future research and a main goal of functional genomics.

Proteomics of *B. subtilis* is even older that the publication of the genome sequence, first proteomic pictures date back to the mid-eighties (1) relying on 2DE, a highly sensitive technique introduced by O'Farrell, Klose and others almost 30 years ago (61,62). Already at that time it was possible to look for changes in protein patterns in response to stress or starvation stimuli at a "proteomic scale", a field mainly pioneered by Fred Neidhardt and Ruth Van Bogelen for *Escherichia coli* (63-66). However, it was almost impossible or at least very difficult to identify the interesting proteins. Later, N-terminal sequencing and the growing DNA databases opened the chance to identify some of the proteins under study already before the entire genome sequence became available. Finally, the publication of the genome sequence was a real breakthrough in proteomics, because it allowed routine identification of proteins on 2-D protein gels by means of mass spectrometry (MS), mainly by MALDI-TOF-MS.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Even if non-gel based alternatives to 2-DE are coming more and more into the focus (69-70). 2-DE is still state-of-the art and will continue to be particularly valuable for bacterial physiology and comparative physiological proteomics involving multiple samples. The high reproducibility of 2-DE is particularly valuable for multi-sample comparisons such as kinetic studies. Non-gel based technologies on the other hand make groups of proteins accessible to proteome analysis that have not been covered thus far, such as kinetic studies.

1.6) THE AIMS OF THIS STUDY

This research aims to study the differential expression of protein during the bacterium *B. stearothermophilus* TLS33 under various stress conditions, for instance, salt stress, cold stress even ethanol stress. Proteomics will be applied in order to gain the basic knowledge which indicates which proteins are responded in individual stress. Furthermore, the post-translational modifications of oxidative stress-induced protein by mass spectrometry approach. In addition, fluorescence dye will be applied for detection of the proteins which were oxidized at thiol group of cysteines.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved