1 INTRODUCTION

1.1 DIVERSITY OF MICROORGANISMS

Biodiversity is the term used to describe the variety of organisms, which includes the number of species and the genetic variation within species. Diversity of the biological world has both aesthetic and practical value. Different species create the subtle and complex beauty of the world and these diverse organisms in our biosphere contribute to the stability and productivity of our planet. It is clear that life is strongly linked to environment. Sudden environmental changes result in disappearance of some life forms while gradual changes in the environment permit both lower and higher forms of life to adapt to modifications in temperature, salinity, acidity, pressure, humidity, and elemental composition of the surrounding medium.

There are billions of bacteria and fungi of great variety in grams of soil. Some have been isolated and characterized while many remain an untapped resource that may provide the key to solving nutritional or disease problems of the future. Preserving diversity of bacteria and fungi is somehow problematic since they are not visible by naked eye. Preserving them may be as important as preserving diversity of plants and animals.

Most commonly, suitable methods for isolation of microorganisms is to spread a sample (for example soil, leaves, roots, and insect legs) on a Petri plate containing nutrients that will enable a wide variety of microorganism to grow, incubate the plates, and then count the different kinds of colonies that grow on the plates. This is a powerful method but still has some limitations due to the conditions chosen.

Although microbial diversity has gained much attention in the last decades, it is known that only 1 – 5 % of all known microbial life forms have been isolated and studied. This estimated percent is based on microbial ecology studies performed using molecular methods such as detection of the small subunit ribosomal RNA (5S rRNA). Instead of isolating new species of microorganisms, researchers prefer to access the gene of interest for their research, clone and express it into *Escherichia coli*. However, one should keep in mind that only about 75 % of microbial genes can

be cloned and expressed in the more commonly used expression systems available to molecular biologists at present.

One of the alarming consequences of the loss of biodiversity is that we cannot guess today what species will contribute to feeding the world or curing diseases tomorrow.

By studying microbial diversity, the life limits can be more easily understood. The changes can be monitored and environmental changes can be predicted.

Microbial communities can be used as model organisms for biological interactions and evolutionary history. From the point of industrial microbiology, studying biodiversity of microorganisms becomes increasingly important. New isolates can produce primary and secondary metabolites, which are important and valuable for biotechnological processes. Further, their proteins, like enzymes may be valuable for biotechnological production. These resources of new genes and proteins, and their conservation, can be important.

From this point of view, one could define biodiversity as the conservation of all kinds of life comprising the microbial gene pool, with the aim of improving human environment and welfare through prudent utilization of the full potential of living resources [1].

1.2 EXTREMOPHILIC MICROORGANISMS

Microbial life is found all over the world. Diverse communities are even found in places of extreme conditions with respect of temperature, salinity, pH, and pressure prevail which are usually considered too harsh to harbour microbial life. The microorganisms adapted to an existence at the edge of life are termed extremophiles. They include members of the Prokaryotes (domains archaea and bacteria) and the eukarya, including algae and protozoa.

Extremophilic microorganisms have extended our knowledge and understanding of fundamental questions such as the origins of life. They have added a new dimension to biodiversity and encouraged additional efforts and interests to explore what is considered the last frontier of life. Extremophiles, such as thermophilic or halophilic microorganisms, can survive under conditions where most “conventional” organisms can not [2, 3]. Usually, extremophilic microorganisms are adapted to biotopes combining several stress factors of the environment which indicate the richness of biological diversity and the possibilities of life to adapt to extreme habitants and to colonize unique ecological niches [4].

Extreme conditions can refer to physical extremes like temperature, pressure or radiation, but also to geochemical extremes such as salinity, and pH. In many cases, they tolerate combinations of more than one environmental stress factor. Some of the extremophiles may be considered as "living fossils" since their environment resembles the conditions that may have existed during the time life arose on earth, more than 3.5 billion years ago. In view of these properties, the extremophilic microorganisms may be considered as model organisms when exploring the possibilities of the existence of extraterrestrial life.

1.3 HALOPHILES

For many years there has been a considerable interest in a group of extremophiles known as halophiles. Halophiles are the indigenous organisms of salt packs, brines, or bodies of salt water. The word halophile is derived from two Greek words "*halos*" for salt and "*philus*" for loving. Halophiles are salt loving organisms and they thrive in relatively high concentrations of salt.

They include mainly prokaryotic and eukaryotic microorganisms with the capacity to balance the osmotic pressure of the environment and resist the denaturing effects of salts. Halophiles are found distributed all over the world in hypersaline environments, such as natural hypersaline brines, coastal, and even deep-sea locations, as well as in artificial salterns.

Halophiles can be classified as slightly, moderately or extremely depending on their requirement for NaCl. Slightly halophilic microorganisms grow optimally at 0.2 – 0.85 M (2-5 %) NaCl; moderately halophilic microorganisms grow optimally at 0.85 – 3.4 M (5-20 %) NaCl; and extremely halophilic microorganisms grow optimally above 3.4 – 5.1 M (20-30 %) NaCl. In contrast, nonhalophiles grow optimally at less than 0.2 M NaCl [5]. In the following part, the characteristics of moderately halophilic microorganisms are explained in detail, since these microorganisms are in the scope of this study.

1.3.1 GENERAL CHARACTERISTICS OF MODERATE HALOPHILES

Kushner stated that at first glance, these organisms are less exciting than the extreme halophiles and solute tolerant microorganisms. They pose quite sufficiently interesting questions, especially those implied by their ability to grow over wide ranges of solute concentrations. Further work on these relatively less studied microorganisms may be expected to bring dividends in form of insight on the relation

of internal and external solute concentrations, and on the cell-associated ions within the cytoplasm.

The soil habitat is inherently heterogeneous. In any part of a soil, there may be different percentages of salt from one region to another. Most isolates require moderately high salt (5 to 10 %) for optimal growth but they are also able to grow at low salinities (0.5 to 2 %). This probably reflects the ecological difference between water and soil environments, as water being relatively homogeneous and of constant concentration whereas the soil showing heterogeneity and affected by factors like rainfall [3]. During the last decade, the extensive studies on hypersaline environments carried out in many geographical areas have permitted the isolation and taxonomic characterization of a large number of moderate halophilic species. Thus, moderate halophiles are represented by several methanogenic archaea as well as strictly anaerobic bacteria. Although some gram-negative species were considered members of different genera (Halomonas, Deleya, Volcaniella, Flavobacterium, Paracoccus, Pseudomonas, Halovibro, or Chromobacterium), phenotypic and phylogenetic data support their close relationship, and they are currently included in the family Halomonadaceae as members of two genera: Halomonas and Chromohalobacter [6, 7]. The common denominator for all moderately halophilic bacteria is their requirement for salt and their ability to tolerate high salt concentrations. Salt requirement and tolerance are highly variable among different species. Moreover, these parameters are by no means constant, since they may vary according to the growth temperature and the nature of the nutrients available [8]. Salt requirement and tolerance may be temperature dependent. In certain halophilic species the minimum and optimum salt concentrations shift to higher values with increasing temperature [9], and similar phenomenon was observed in halophilic bacteria as well.

For halophilic microorganisms, the important issue is the achievement of a high osmotic pressure in the cytoplasm while keeping the Na^+ concentration low. One strategy involves accumulation of K^+ and Cl^- ions to maintain osmotic balance. This mechanism is used by a number of halophiles. The aerobic halophilic archaea of the order halobacteriales accumulate KCl at concentrations at least as high as the NaCl concentration in their surrounding medium. This type of osmotic adaptation costs relatively little energy. The second strategy of osmotic adaptation is to exclude salts from the cytoplasm as much as possible with transport mechanisms generally based on Na^+ /H^+ antiporters [10] and to accumulate organic solutes to provide osmotic balance. High osmolarity in hypersaline conditions can be deleterious to cells since water is lost to the external medium until osmotic equilibrium is achieved. To

prevent loss of cellular water under these circumstances, halophiles generally accumulate high solute concentrations within the cytoplasm [11]. A variety of compounds is used for this purpose, ranging from glycerol and other sugar alcohols, amino acids and derivatives such as glycine betaine and ectoine (2-methyl-1,4,5,6-tetrahydropyrimidine-4-carboxylic acid) and its 5-hydroxy derivative, to simple sugars such as sucrose and trehaose [12] which do not disrupt metabolic processes and have no net charge at physiological pH. This strategy is widely used in all three domains of life. The use of organic osmotic solutes requires much less far-reaching adaptations of the intracellular enzymatic machinery then does the accumulation of KCl. However, the production of such solutes is energetically expensive [13].

Studies based on 16S rRNA sequence analysis have permitted a determination of the phylogenetic position of most moderately halophilic bacteria. The levels of 16S rRNA sequence similarity among these species ranged from 91.5 to 100 %. Although several sub groups, which might represent separate genera, were resolved, that could not be differentiated on the basis of phenotypic or chemotaxonomic features. The 16S rRNA sequence analysis supports the placement of these species in a separate genus [14]. In the literature it is clearly seen that most of the halophilic organisms for which the 16S rRNA sequence analysis was done, belong to the archaeal groups rather than to the moderately halophilic bacteria. According to the code of nomenclature of bacteria, several other moderately halophilic aerobic bacteria exist which are not yet characterized taxonomically.

Besides 16S rRNA sequence analysis, strongly conserved 5S rRNA subunits were also used as phylogenetic markers. Nucleotide sequences of microorganisms belonging to different taxonomic groups including halophilic microorganisms are collected in '5S Ribosomal RNA Database', which is updated systematically [15-18]. This database contains updated taxonomic classification of microorganisms including 16 halobacteriales with individual sequence records of 5S RNA given at each taxonomic group (http://biobases.ibch.poznan.pl/5SData/).

1.3.2 BIOCHEMICAL CHARACTERIZATION OF HALOPHILES

1.3.2.1 ION CONCENTRATIONS INSIDE THE CELL

To cope with the high and often changing salinity of their environment, the aerobic halophilic bacteria, similar to all other microorganisms, need to balance their cytoplasm with the osmotic pressure exerted by the external medium. Osmotic balance can be achieved by the accumulation of salts, organic molecules, or a

combination thereof. Another possibility is that the cell is able to control water movement in and out and maintain a hypoosmotic state of their intracellular space [19-21]. In many *Halomonas* species the sum of the halotolerant intracellular Na^+ and K^+ concentrations is much lower than the medium concentration [22]. The most important and studied intracellular ions are sodium, potassium and chloride. In moderate halophiles, the apparent intracellular Na^+ concentrations are often too high to enable the generally salt-sensitive cytoplasmic enzymes to function. In addition, Na^+ and other ions may be bound to the outer cell layers, in amounts increasing with external salinity. For potassium, it is recorded that the accumulation is up to a few tenths of 1 M [23]. In contrast to some archaeal halophiles, in moderately halophiles, cytoplasmic potassium contributes relatively little to the achievement of an osmotic balance. Estimations of intracellular chloride concentrations within the cells of moderately halophilic bacteria are greatly variable, from relatively low values of 55 and 139 mM to values as high as 0.7 to 1 M [24]. The measured intracellular Cl^- concentrations are in most cases much lower than combined Na^+ and K^+ concentrations [23]. When cells are grown at higher salt concentrations, the apparent intracellular Cl^- concentration may be increased [25]. The old assumption that Cl^- is the main counter ion for the intracellular cations in moderate halophiles is not necessarily true, and since no other anions have been detected at high concentrations within the cells, it has been speculated that most of the cellular cations may be associated with the negative charges present on proteins, cell envelopes, and other macromolecules [26, 27].

It has often been speculated whether the exclusion of salt and the maintenance of ionic gradient by the moderately halophilic bacteria is achieved through the constant pumping of ions by active, energy-dependent mechanisms or by tightening of the permeability barrier of the membrane. Continuous-culture experiments indicate that the maintenance energy of these organisms is relatively independent of the salinity of the medium, providing evidence in favour of an ion-tight membrane [12].

A low intracellular ionic environment maintaining the intracellular salt concentration at a level considerably below that of the environment can be achieved only by energy-dependent mechanisms [28]. To achieve low intracellular Na^+ ion concentrations against a constant influx of Na^+ ions leaking inside through a not completely impermeable membrane, in addition to Na^+ entering the cells during co-transport with amino acids and other substrates, mechanisms of Na^+ extrusion have to be present in the cell membrane. Two possible mechanisms have been suggested: activity of Na^+/H^+ antiport and presence of primary respiration- driven Na^+ pump. It is generally accepted that Na^+/H^+ antiport activity is an important mechanism for

maintaining low intracellular sodium concentrations. However, there is still considerable controversy about the occurrence and relative importance of primary Na^+ pumps in different representatives of the very heterogeneous group of moderate halophiles.

1.3.2.2 ORGANIC OSMOTIC SOLUTES

Since the intracellular ion concentrations measured in moderate halophiles are generally insufficient to provide osmotic balance with the external medium, organic compounds accumulated by the cells have been searched and only relatively recently the principle organic osmotic solutes of the moderately halophilic bacteria have been identified. Moderate halophiles accumulate low-molecular-weight organic compounds in their cytoplasm to cope with the osmotic stress and to maintain positive turgor pressure without interfering with the metabolic functions of the cells.

Compatible solutes are polar, highly soluble molecules and are uncharged or in zwitterionic form at physiological pH. They are strong water structure formers and effective stabilizers. Water near the interfaces is structurally different (more dense), allowing the osmotic solutes to migrate toward the less dense water fraction. This phenomenon of non-specific exclusion is often described in terms of increased surface tension of water, with the presence of solutes affecting the forces of cohesion between water, minimization of entropy, and reinforcement of the hydrophobic effect. Compatible solutes display a general stabilizing effect by preventing the unfolding and denaturation of proteins caused by heating, freezing and drying [11, 12, 29, 30, and 31]. Amino acids, glycine betaine, ectoine and hydroxyectoine are the best known and dominant organic solutes known.

The ability to produce and accumulate high concentrations of these compounds makes moderate halophiles useful for the biotechnological production of these osmolytes.

Before the importance of these compounds was realized, unsuccessful attempts to identify intracellular organic compounds at high concentrations led to the view that these bacteria may be able to manage without the traditional osmotic balance and may maintain a hypoosmotic cytoplasm [20]. The discovery that ectoine and hydroxyectoine, two compounds that have long resisted detection, may be present inside the cells in molar concentrations abolished the need to hypothesise a hypo osmotic intracellular space [32].

Ectoine and hydroxyectoine exert a remarkable protective effect on a number of labile enzymes such as lactate dehydrogenase and phosphofructokinase. In a study with several compatible solutes (glycine betaine, trehalose, glycerol, praline, ectoines, and sugars), hydroxyectoine showed the highest efficiency for the protection of lactate dehydrogenase against freeze thaw treatment and heat stress whereas ectoine was the most effective freeze-stabilizing agent for phosphofructokinase [11, 33]. Ectoine and hydroxyectoine are found in high concentrations in different *Halomonas* species and are the dominant sources in cells lacking glycine betaine or its precursor, choline [32, 34].

1.3.2.3 AMINO ACID COMPOSITION OF HALOPHILIC PROTEINS

The aerobic halophilic bacteria tolerate high salt concentrations providing osmotic balance with the organic solutes. However, the apparent intracellular salt concentrations can be fairly high. Therefore, it could be expected that extracellular and membrane bound proteins of the halophilic bacteria may display halophilic characteristics in amino acid composition similar to those found in the halophilic archaea keeping in mind the buffering effect of organic solutes.

Environmental peculiarities of halophilic microorganisms are expected to require protein features clearly distinguished from those of proteins from microorganisms living in the "normal" environment. The halophilic proteins bind significant amounts of salt and water, in solvent conditions, which are similar to their physiological environment. This characteristic seems to be in contrast to that of non-halophilic proteins, which bind similar amounts of water but do not bind, salt. A number of studies have suggested that the halophilic adaptation correlates with an increase in acidic amino acids in the protein composition [35-38]. The fraction of acidic residues is extremely large in the surface composition of the halophilic proteins. The large number of acidic residues on the surface of halophilic proteins has been rationalized on the basis of their superior water-binding abilities in the charged forms [39]. An increase in salt concentration from 1 M to 5 M increases the pKa value of aspartic acid from 4.0 to 4.9, and that of glutamic acid from 4.4 to 5.3 [40], and decreases the population of the charged forms of these acidic residues at the neutral pH.

1.3.2.4 CELL ENVELOPES OF MODERATE HALOPHILES

The cytoplasmic membrane forms the barrier between the cytoplasm (generally low in salt) and the environment with its high salinity. The properties of these membranes are regulated by the outside salt concentration to adjust such functions as ion permeability and the activity of integral membrane proteins [41]. The envelopes of

the moderate halophiles are involved in regulating the cytoplasmic ionic environment by both binding and pumping ions, and halo adaptation is at least in part a response of the cell envelope to osmotic stress [42].

Studies of the effect of salinity on polar lipid composition of the membranes have consistently shown that the higher the salinity, the higher the content of negatively charged phospholipids at the expense of neutral phospholipids. With increased salt concentration, the amount of phosphatidylethanolamine (PE, uncharged) is decreased and there is an increase in the amounts of the negatively charged phosphatidylglycerol (PG) and/or diphosphatidylglycerol (cardiolipin, CL) [20, 41, 43]. To explain the shift toward a higher negative-charge density on the membrane at increasing salt concentration, it was first postulated that the increase in the amount of anionic lipids serves to allow charge balance at the membrane surface exposed to high Na^+ concentrations [44, 45]. However, a simple calculation showed that millimolar concentrations of salt suffice to provide the negative charge shielding [42]. Moreover, a similar increase in the amount of negatively charged polar lipids was observed when non-ionic solutes were added to the medium to increase its osmotic value. Another idea proposed was that the high content of negative phospholipids might contribute to the regulation of the selective permeability of the membrane to cations [46]. It is now assumed that the charge in polar lipid composition provides a mechanism for preserving the membrane bilayer structure. PE containing unsaturated fatty acids tends to form non-bilayer phases, while PG forms bilayers. Addition of PG to PE in the appropriate amounts could counteract the effect of increased salinity. A functional membrane requires a suitable proportion of bilayer-forming and non-bilayer-forming lipids.

1.4 PROTEOMICS

The knowledge of any genome and its polymorphisms are the basis for an understanding of the gene functions in cellular processes. However, the data of the genes yields information only about the static state of the inherited information. In order to understand the dynamic processes within the cells, e.g. during cell differentiation and regulation or under disease development, it is mandatory to study the proteins involved. Their presence, actual concentrations and possible modifications must be known in order to gain clues about their functional activities. This means that entire protein extracts of cells or tissues have to be analysed in order to learn about the expression rates of the proteins and their various species under certain conditions, in certain cells, tissues or organs. These analyses are known as studies in "**Proteomics",** an abbreviation of the words **Prote**in and Gen**omics**, firstly

proposed 1994 by M.R. Wilkins in Sydney, Australia [47]. The proteome differs from tissue to tissue, from early stages of disease development to later phases, and shows differences in protein expression rates also under medical treatment. Each cell of an organism contains the same identical genome but many different proteomes.

The main differences between genomic and proteomic studies hence are:

1. In opposite to genome analysis the Proteome is characterized by a dynamic event and it reflects the functional status of the biological system. Any stress factors affecting it, can cause a change at the proteomic level.

2. Protein expression may be different from mRNA turnover and stability, which causes differences between the level of expressed proteins and the number of mRNAs in a cell, at a specified condition [48].

3. Post translational modifications of proteins largely affect their function, and this can not be evaluated on the genomic scale.

1.4.1 PROTEOMICS IN STUDYING MICROORGANISMS

Proteomics is widely used to analyse the bacterial protein content. The identification of proteins and protein expression patterns under given physiological conditions by proteomic analysis has gained fundamental importance for functional studies of cellular processes in recent years. All these studies help us to collect data about microorganisms and their cell processes.

Collecting all genomic, proteomic and bioinformatic data of a specific organism enables us to produce a global look of their specific and important features [49].

Proteome studies of known microorganism have been reviewed by several research papers. All these studies are the collection of different types of information like comparison analysis of bacterial, archaeal and eukaryotic proteome and genome [50] technological developments in studying bacterial proteomes [51], mass spectrometric approaches to bacterial proteome studies [52] and, *in silico* studies of bacterial proteomes [53]. The information collected from different studies that contribute to bacterial proteomics, enables knowledge flow about bacterial proteins responsible for structure, growth, division, and energy production. Up to date, complete genomes of 28 archaea, 370 bacteria and 17 incomplete genomes have been reported as listed in the databases (http://www. tigr.org).

1.4.2 TECHNIQUES USED IN BACTERIAL PROTEOME STUDIES

Proteomics, whether we study a specific pathway or study molecular networks, has main steps to follow, which are the extraction, separation, detection of proteins, the identification of proteins of special interest and the data analysis under specified conditions of growth, isolation, cell disruption and type of analysis.

1.4.2.1 EXTRACTION AND SEPARATION OF PROTEINS

1.4.2.1.1 EXTRACTION

The extraction method, which might be used for bacterial studies, should be chosen very carefully in order not to loose or destruct any proteins that could be interested. Since the aim is to keep all the proteins of interest in a soluble fraction, one should first decide how to lyse the cell without causing harm to the content. Surely, both the choice of the disruption method and lysis buffer is very important at this stage.

For bacterial cell disruption, ultrasonication and, freeze–thaw are two popular methods used.

For ultrasonication, frequencies above 20 kHz at given temperatures are applied to the solution, and gaseous cavitations occur. As the gas bubbles collapse, shock waves are formed and this causes the disruption of the cells in that solution. This method is successfully employed for microorganisms of any kind. The main limitations are temperature and detergents used in lysis buffer. There is a high generation of heat during sonication. To overcome this problem, the cells are kept on ice and sonication is done in turns of 30 seconds for several cycles. The bubbles created by the detergents cause sample loss in the disruption step, for which should also be taken care.

Another technique used for cell disruption is freezing and thawing. Repeated freezing and thawing cause an intracellular ice formation, which results in extensive membrane lesions whereby periplasmic and intracellular proteins are released completely. In this method the freeze temperature, time period of each freeze-thaw cycle, and the thawing temperature are important factors. The shock freezing is supplied by liquid nitrogen. Sudden freezing and thawing is repeated for about 1 hour totally with about 4 times cycling.

There are several lysis buffers and commercial extraction kits recommended for different purposes of bacterial protein extraction. The choice of a lysis buffer depends on the sample to be lysed and also which protein fraction to be needed for

the analysis. Special lysis buffers can be used for isolation of membrane protein fraction, cytosolic protein fraction, or whole cell protein fraction [54 - 57]. For both, commercially available kits and for home made lysis buffers, one should be careful to include some additional chemicals to minimize the interference effects of nucleic acids, lipids, salts, proteases and polysaccharides. These are potential interfering substances that disturb either protein solubilisation or electrophoretic protein separation.

DNA and RNA molecules are composed of nucleic acids and these molecules can bind proteins through electrostatic interactions causing an increase in sample viscosity and therefore obstruct polyacrylamide pores. In order to lower this effect of nucleic acids, nucleases like RNAse and DNAse, or ampholytes are included in the lysis buffer.

Lipids can also bind proteins causing artificial migrations and streaking. To avoid this effect, non-ionic detergents like CHAPS (3-[(cholamidopropyl) dimethylammonio]-1-propane sulfonate), or Triton X-100 are used in the lysis buffers.

Salts have bad effects on electrophoresis systems. They can give rise to excessive local differences of heating due to high current. Bacterial growth medium can contain high percentages of salt. Removal of salt is done by washings of the cells, that of intracellular excess salt is more challenging. The common method is to use centrifugation to remove salt. But it should be kept in mind that some samples can be lost in this step, e.g. membrane proteins may stay in the sediment.

Following the lysis, proteases can directly get activated and generate artificial spots of breakdown products. To avoid this, protease inhibitors are added prior to cell disruption. PMSF, benzamidine, and leupeptine can be added to inhibit serine and cycteine proteases and EDTA for metalloproteases.

To avoid the side effects of polysaccharides, high-speed centrifugation is preferred. However, some polysaccharides like mucins are much more difficult to remove. Addition of tributyl phosphine and thiourea are employed [58] to remove such polysaccharides.

Phosphate and tris buffers are most commonly used buffers for bacterial protein extractions. If a native lysate is needed where the proteins function and structure is preserved, then addition of denaturing agents should be avoided. It can be meaningful to prepare the lysates with appropriate buffer including inhibitors of

several interfering agents discussed above without denaturants and reserve part of it for functional analysis studies and prepare the rest of the sample for two-dimensional electrophoresis where some more chemicals are added to the buffer system to obtain denatured lysates where proteins are fully unfolded, and non-covalent interactions are disrupted, and macromolecular interactions are broken. The denaturants are usually urea at high molar concentrations, e.g. 9 M, and urea/thiourea mixtures of 7 M and 2 M, respectively or some other combinations. DTT (dithiothreitol) is also used at concentrations of 50 – 70 mM for disulfide bridges disruption.

1.4.2.1.2 SEPARATION

In gel bacterial proteomics, the most appropriate method to be used is high-resolution two-dimensional electrophoresis (2-DE) for separation of bacterial proteins where thousands of proteins can be separated at once [59]. In 2-DE technique, the first dimensional separation is done depending on the isoelectric points of the proteins, and the second dimension according to the molecular weights of the proteins. Separating proteins according to these two physico-chemical properties enables thousands of proteins to be differentiated from each other as the proteins appear as individual spots on the polyacrylamide gels.

The isoelectric separation can be made in different ways, one technique using immobilized IPG strips [60], alternatively, the second way is the home made gels casted in capillaries applying the IEF-NEPHGE (Non-equilibrium pH gradient electrophoresis) method [61, 62].

Immobilline IPG strips, the so called “ready to use” gels, which contain immobilized polymers with carrier ampholytes of different charges to set up a pH-gradient in the strips, e.g. between pH 3 and 10, can be supplied from many companies and are popular among many research groups. This technique is discussed in detail in several books, papers and reviews [60, 63, and 64]. However, this IEF-technique possesses severe problems. The sample-loading step is problematically since the protein mixture is soaked into the strips during rehydration, which takes several hours. Therefore precipitation of proteins can occur which are lost for the further separation. Furthermore, in the 2-DE-gels at maximum 2.000 proteins can well be separated. Often the proteins occur with streaks and stripes so that their identification is troublesome. In total protein extracts from cells or tissues, which contain up to 30.000 proteins, often several proteins occur per spot so that clear protein identification can not be made.

In the NEPHGE technique the pH scale is built up with soluble carrier ampholytes of appropriate mixtures of different pH compounds. During the run, proteins are not focused completely to their isoelectric point (so precipitation of proteins is prevented), but instead move at different rates across the gel. For this reason, the accumulated volt hours actually determine the pattern spread across the gel. It is therefore imperative that volt hours be consistent to assure reproducible patterns. The protein samples are loaded onto the acidic side of the gel and separated in a pH gradient, e.g. fro 3 to 10. Together with the second dimension (see below) this technique is a high – resolution technique that allows resolution of 5.000-10.000 protein spots depending of the size of the gel as reported [64]. The advantages of using this technique is the freedom to use different sizes of gels in glass capillary rods which allows giant gels like 48 cm in length [64]. Also the ability to make thicker and thinner gels saves samples to be loaded onto the gels. Analytical gels are prepared in 0.9 mm i.d. glass capillary rods at a length of 24 cm and preparative ones are prepared in 1.7 mm i.d. glass capillary rods at a length of 24 cm. The NEPHGE technique is a sophisticated method for separating a high number of proteins and is applied in experienced laboratories.

The second dimension of 2-DE usually is made by SDS-PAGE (sodium dodecyl sulphate- polyacrylamide gel electrophoresis). The gels that run at first by the NEPHGE method are re-equilibrated for the second dimension buffer system and placed on top of the second dimension gel which can be 10-12% acrylamide or a gradient separation gel. It should be taken into account that, working with gradient gels at this stage, can cause reproducibility problems. The length of the NEPHGE gel (e.g. 24 to 30 cm) is the width of the SDS-PAGE gel showing the pH scale at the x-axis of the 2-DE gel. The height of the SDS-PAGE gel gives the molecular weight scale (y- axis of a 2-DE gel). After the secondary run (SDS-PAGE), the proteins have to be stained and the map of protein spots yields information about their physicochemical characteristics. 2-DE by the NEPHGE technique has many advantages, but special organisms like halophiles; the conditions must be optimised to obtain well-separated and reproducible data.

1.4.2.2 DETECTION OF PROTEINS

In 2-DE, as stated above, the second dimension usually is SDS-PAGE and detection methods developed for normal SDS-PAGE can also be applied for 2-DE gels. The important parameter of choosing a good detection method depends on the further analysis of the protein spots. Organic dyes, silver staining and, fluorescence staining

are the most commonly used staining methods. Another alternative of detecting proteins is electroblotting.

Among the organic dyes anionic Brilliant Blue G and R dyes are commonly applied ones, which are also known as Coomassie Blue G and R. They have high affinity for proteins resulting in good detection limits about 50 ng. These anionic dyes have advantages in acidic media, e.g. when fixing proteins on gels in acetic acid buffers. A method with a higher sensitivity is developed for Coomassie Blue G dyes [65, 66]. Coomassie Brilliant Blue G-250 forms miroprecipitates in acidic media containing ammonium sulphate and this makes colloidal staining possible. Since there is a very little amount of free dye in the medium the background staining is minor and easily removed by washing steps. This staining protocol allows keeping the gels in the dye solution for several days to achieve a satisfying staining.

When the sensitivity limits of Coomassie Brilliant Blue dyes are compared, the highest sensitivity is reached with Coomassie Brilliant Blue G- 250 colloidal staining with a detection limit of nanograms; the least sensitive stain is by Coomassie Brilliant Blue R-250 with a detection limit of micrograms. The Coomassie stains allow Edman sequencing and MS analysis of the spots, which makes them important for proteomics.

Another type of staining method popular in proteomics is the silver staining firstly introduced in 1979 [67]. Since then, several modified and improved procedures to achieve better sensitivity and lower background were recommended. The driving force for staining is the affinity of proteins for the cation, which is silver in this case [68]. In the electrophoresis media, many molecules like SDS, chloride and amino acids show affinity for silver so it is important to remove all these interfering substances from the media for low background. Silver ions bind to the proteins amino acids side chains mostly to sulfhydryl and carboxyl groups [67, 69, 70] followed by reduction of free metallic silver [71, 72]. Each of different silver staining protocol has its own advantage regarding timing, sensitivity, cost and compatibility with analytical methods, especially MS.

The protocols using glutaraldehyde or formaldehyde which causes the cross linking of two lysine residues within the protein chain results in limitations of the trypsin digestion and protein extraction losses [71]. Several modifications of silver nitrate staining procedures have been developed excluding glutaraldehyde and formaldehyde; which are compatible with analytical methods like sequencing or mass analysis [73, 74]. The sensitivity reached by silver staining protocols is in the few nanogram levels, which allows detection of low abundant proteins. Clear

background and very high sensitivity of silver staining makes this method favourable for computerised analysis.

Another sensitive method of visualizing proteins of 2-DE gels is fluorescent staining at a sensitivity level of one nanogram. From the two main different fluorescent methods the first utilizes a fluorophore covalently coupled to proteins at the pre-electrophoresis step. This is known as a labelling step but complicated by protein modification at certain pI points. To overcome this unwanted pI modification of the proteins mostly neutral, thiol reactive probes are used for labelling [75]. In the second method, fluorescent detection is provided by non-covalent, environment sensitive probes. The probes used in this staining procedures are weakly fluorescent in water but highly fluorescent in apolar media like detergents. In case of electrophoresis this media would contain SDS as a detergent and at the binding side of the detergent to protein a fluorescence-promoting environment is achieved. Sypro Dyes from molecular probes [76, 77] or Nile Red [78] are the most common used dyes. SYPRO stains are more sensitive than the standard Coomassie stain and as sensitive as silver stain. The fluorescence intensity of SYPRO dye stained spots are linear with protein quantity over three orders of magnitude, a much broader range than either CBB or silver staining [76]. The staining protocols are easy and can be completed within an hour and they are compatible with most of the microanalysis. On the other hand it should be remembered that it implies a rather delicate hardware since the emitted light is usually very low in intensity. The dyes are expensive and the image of the fluorescence stained gel should be obtained by a fluorescence imager, which is also rather expensive.

Electroblotting means transfer of electrophoretically separated proteins from the polyacrylamide matrix onto a membrane, which is an adsorbing surface for proteins, at an electric field. This technique is described by Renart and by Towbin individually at the same year [79, 80]. Electroblotting, in combination with 2-DE is one of the most versatile method used in microsequencing and / or mass spectrometry. Blotting on membranes possess some advantages like easy handling, easy storage until analysis and, providing suitable environment for microdetection of proteins, and contamination free media like a minimum of salt and detergent.

There are three major methods used to transfer proteins electrophoretically to a membrane, semidry blotting, tank transfer blotting and diffusion blotting. Although tank transfer blotting wastes large volumes of buffers it is still very widely used. On the other hand, it seems to possess no advantage over semi-dry blotting. Transfer by diffusion blotting, which is analogous to the original procedure developed by

southern for DNA, is possible. However, because of the highly cross-linked nature of the acrylamide gels used for protein separations this method can last over days.

Semidry blotting has become increasingly popular as it is fast, uses little buffer, and the progress of the transfer can be readily checked. This system is firstly described by Kyhse & Andersen in 1984 [80, 81]. This system consists of two plates as electrodes in which a blot sandwich is made horizontally. The sandwich contains layers of filter paper soaked into the transferring buffer, a wetted membrane on top of the filter papers, gel to be transferred and another group of filter paper layers of the same type, successively arranged from the anodic side.

The efficiency of electroblotting is directly correlated with the efficiency of transfer of protein spots to the membrane and also that of adsorption of protein onto the membrane surfaces and influenced by some parameters as the transferring buffer systems, SDS, methanol, and membrane selection.

A good buffer should efficiently elute the proteins out of the gel and deposit them in high yield on the membrane. A wide variety of buffers have been used in literature [80, 82-84]. For Na-borate buffer systems, it was found that the molarities less than 10 mM and above 100 mM cause decreased transferred yields [85]. The pH between the gel and the membrane is arranged around 8 due to the electrophoresis buffer remaining in the gel matrix. In a good transferring buffer a limited amount of SDS is used to provide a good transfer of hydrophobic and high molecular weight proteins. With an addition of methanol a significant increase of transfer yields is obtained. The stability of the SDS-protein complex is affected by methanol, which dissociates the complex more easily. Further, the hydrophobic, uncharged membranes need the presence of methanol for the interaction between the membrane surface and the protein. The blotting is further influenced by the choice of the membrane. Several membranes are available through several companies, which have different composition, and textures like glass fibre based, modified with some positively charged groups, by hydrophobic groups or, are pure organic polymers like polyvinylidene difluoride (PVDF). The selection of a membrane suitable for the further analysis is important at this step. During membrane selection, both the chemical and the physical characteristics should be taken into consideration.

Staining methods for blotted membranes after 2-DE gels depend on the further analysis: Immunostaining is one of the most important staining methods for visualizing the antigenic proteins. For microsequencing and mass spectral analysis, Coomassie Blue staining of membranes is very popular. Colloidal gold stain, which

is an irreversible stain, can be applied to membranes for total staining of blotted proteins.

1.4.2.3 IDENTIFICATION METHODS FOR 2-DE GEL PROTEINS

In a 2-DE gel, depending on the origin of the sample, hundreds or even thousands of protein spots are separated at once. Blots of these gels with a good electroblotting step can be nearly fully transferred on to the appropriate membrane. When the interesting spots are visualized, generally only few picomoles of the proteins are present but in some lucky cases, hundreds of picomoles can be present in one spot. This situation affects the method to be chosen for identification. The parameters that effect the selection of the appropriate identification are the amount of the protein, the chemistry to be used, the origin of the sample and the databases available for that sample. There are several methods like protein partial hydrolysis coupled with chromatography, N- terminal sequence analysis by Edman degradation, ladder sequencing coupled with mass spectrometry, and mass spectrometry coupled with database analysis.

1.4.2.3.1 N-TERMINAL EDMAN DEGRADATION

In 1949 in Melbourne, Australia, the Swedish scientist Pehr Edman developed a method for sequential degradation of peptides and proteins known as “Edman degradation” [86] and detailed the chemistry in his following reports [87, 88]. In 1967, he first developed an automated protein sequenator which became more sensitive than any of the available manual sequencing methods but still needed about 7 mg amounts of purified protein [89]. In 1971, the American company Beckman copied his invention and sold the machine as Beckman sequencer. Later in 1981, Applied Biosystems, a new company at Palo Alto, CA, sold a more sensitive sequencer for Picomole amounts of proteins equipped with the more sensitive cartridge-sequencing device.

The N-terminal sequence of a protein or a peptide can be determined by the automated Edman degradation, which is still one of the most powerful methods for protein studies. In the most recent versions of a fully automated system, a few picomole quantities of proteins can be N-terminally sequenced. With a successful N-terminal analysis, up to 50 residues of a protein and 10 or 15 amino acids of a peptide can be sequenced which is sufficient for protein identification by comparison with all protein sequences in the databases on internet. Although it is expensive and impossible when there is a blockage at the N-terminal residue (which can happen

often), it is still favourable since it is sensitive, reliable and generates '*de novo*' sequences of peptides.

By Edman degradation, due to the reaction with phenylisothiocyanate (PITC) the N-terminal amino acid of a peptide or a protein backbone is cleaved off as an anilino-thioazolinone (ATZ-amino acid), which is extracted and converted to the phenylthiohydantoin (PTH-amino acid derivative). The latter is analysed by HPLC in comparison to a mixture of all synthesized PTH-amino acids. In this manner, by repetitive reactions, the amino acid sequence of a peptide or protein is determined. Accordingly, Edman degradation involves 3 main steps, which are the coupling, the cleavage and the conversion. After conversion, the derivatized amino acid is identified.

1.4.2.3.2 MASS SPECTROMETRY

The basic principle of mass spectrometry (MS) is to generate ions from compounds by any suitable method, to separate these ions by their mass-to-charge ratio (m/z) and to detect them qualitatively and quantitatively by their respective m/z and abundance [90].

Mass spectrometry, which is used for identification of proteins and peptides, now is a popular method since it has the ability to solve structural problems quickly and with pico- to femtomole amounts. There has been a huge improvement through the years on both technical and instrumental developments and this is still continuing yielding precise results with very accurate mass determinations. Due to the high sensitivity of modern mass analysers, 50 femtomoles may identify the analysed protein or peptide. This property makes this method popular for identification.

All mass spectrometers have three major components, which are required for measuring the masses of molecules that have been ionised in the gaseous phase. These components are the ion source, the mass analyser and the detector. Besides these 3 main components, the instruments are equipped with sample inlet and data system. Additionally they may be attached to an HPLC-system (LC-MS), and may contain an ion trap.

The mass instruments consist of different ionisation methods, such as electron ionisation (EI), chemical ionisation (CI), field ionisation (FI), fast atom bombardment ionisation (FAB), matrix-assisted laser desorption ionisation (MALDI), or electrospray ionisation (ESI), and they have mass analysers, such as Time-Of-Flight (TOF) instruments, linear quadrupole instruments (LQ), three

dimensional quadrupole instruments (3DQ), Fourier transform ion cyclotron resonance instruments (FT-ICR). Besides these there are several hybrid mass spectrometers that combine different types of ion sources and/or mass analysers in a single instrument.

For Protein studies, most commonly used mass spectrometers are MALDI-TOF MS, ESI-MS/MS, LC/ESI-MS/MS, Nano- ESI/Q-TRAP, ESI-Q-TOF MS. Currently, QqTOF systems are the most successful hybrid instruments used for protein studies.

1.4.2.4 IDENTIFICATION OF PROTEINS BY ELECTROSPRAY IONISATION TANDEM MASS SPECTROMETRY (ESI-MS/MS) AND DATA ANALYSIS

ESI-MS/MS is based on producing molecular ions of peptides or peptide mixtures by an electrospray-technique (electrospray-ionisation). By using ESI, these ions are produced directly from liquids at atmospheric pressure and injected into the pre-vacuum of the ionic source to withdraw the solvent. Enzymatically digested protein samples (trypsin is commonly used) are dissolved or prepared in appropriate solvent mixture (encouraging ion formation) containing methanol, water, formic acid and acetonitrile. The solubilised sample is then infused into a capillary at a constant flow rate entering the ionic source where the intact molecules are ionised in the gas phase, free of solvent and other molecules. The ions are individually selected and separated according to the system used.

1.4.2.4.1 IONISATION

1.4.2.4.1.1 Ionisation by Electrospray

Electrospray ionisation (ESI) is a soft ionisation technique used to transfer ions from solution to the gas phase. This ionisation method is successfully applied to large, non-volatile, chargeable molecules like peptides, proteins, nucleic acids and polymers [91]. A spray of fine, highly charged droplets is created at atmospheric pressure in the presence of a strong electric field. The ESI source can be just a metal capillary at elevated voltage relative to a counter electrode (interface plate) with an orifice where the ions entrained in a flow of gas enter the mass spectrometer. Liquid flow is supplied by syringes, HPLC, capillary electrophoresis with flow rates of one to a few microliters per minute. Ionisation at lower flow rates as nanoliter can be achieved by “nano LC” systems [92]. Wilm and Mann demonstrated that a more narrow spray capillary, which is a borosilicate glass, results in smaller droplets and much more reduced flow rates. NanoESI allows for high polarity solvents such as

pure water in both negative and positive ion mode, and has extremely low sample consumption [93] and tolerates higher buffer salts [94, 95]. These characteristics of the nano system enable to work with very small amounts of samples, generally the case working with 2-DE based proteomics (Figure 1.1).

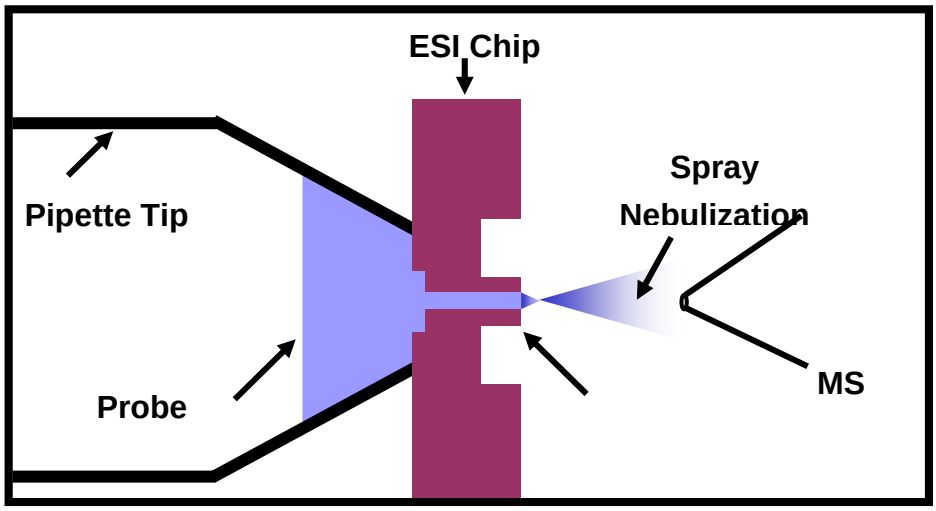

Figure 1.1 Electrospray ionisation

1.4.2.4.1.2 Ionisation by Matrix Assisted Laser Desorption

The term matrix-assisted laser desorption ionisation (MALDI) was coined in 1985 by Franz Hillenkamp, Michael Karas and their colleagues [96]. This soft ionisation technique is one of the most popular one, which allows the analysis of proteins, peptides and large organic molecules like polymers tending to be fragile and fragment when ionised by more conventional ionisation methods.

Samples to be analysed are mixed well with a matrix, which is used to provide a better ionisation by laser beam and to facilitate vaporization and of biomolecule fragments. The matrix molecules are designed to be fairly low molecular weight allowing vaporization but on the other hand large enough to stay stabilized during sample preparation. The matrix molecules also carry some more important characteristics like being acidic therefore acting like a proton source encouraging the ionisation of the analyte, having strong optical absorption in the UV so that they easily and efficiently absorb the laser radiation and chemically they are engineered with polar groups to be able to function in aqueous environments. The most commonly used matrix crystals are α-cyano-4-hydroxycinnamic acid (CHCA) 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) and, 2,5-dihydroxybenzoic acid (DHB).

When the matrix solution is mixed with the protein sample, the organic solvents allow hydrophobic molecules to dissolve into the solution while water allows hydrophilic molecules to dissolve. The sample solution prepared is spotted onto a special MALDI plate and when the solvents vaporize, the analyte mixed with the matrix is spread throughout the crystals. The laser is fired at the crystals in the MALDI spot cell on the target plate. By absorbing the laser energy the matrix molecule ionises and transfers a part of its charge to the analyte molecule (peptide fragments of proteins), which cause ionisation of the analyte molecules. The ions produced in MALDI ionisation systems are generally singly-charged ions ($[M+H]^+$ in the case of an added proton, $[M+Na]^+$ in the case of an added sodium ion, $[M-H]^-$ in the case of a removed proton) but according to the voltages used and also due to the characteristics of the matrix molecules, multiply charged ions ($[M+nH]^{n+}$) can also be seen (Figure 1.2).

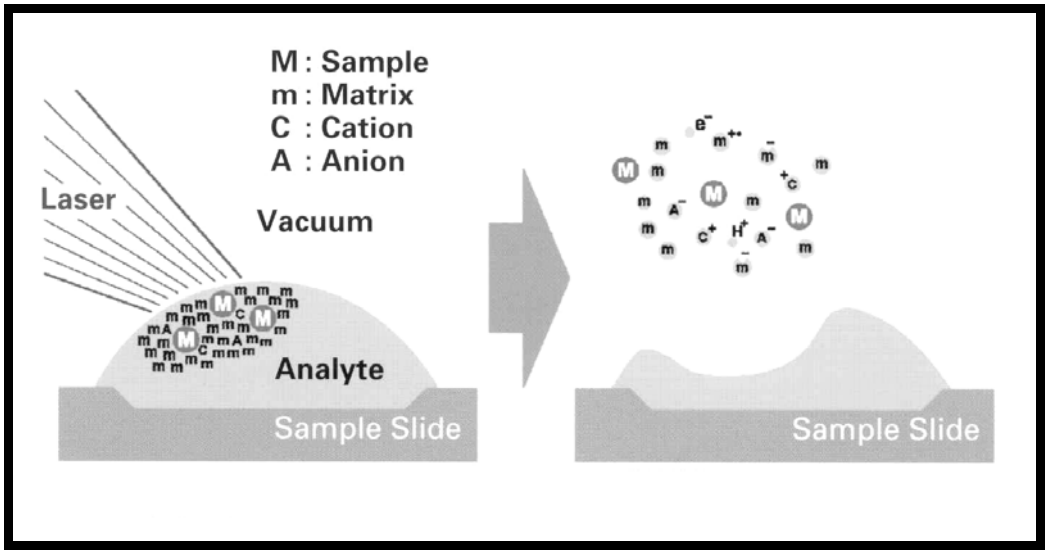

Figure 1.2 MALDI system

1.4.2.4.2 QUADRUPOLE (Q) MASS ANALYSER

The electrospray technique was first attached to quadrupole mass spectrometers and still in use routinely in ESI-MS [97, 98]. For protein identification the triple quadrupole (TQ) MS is superior as a device for the generation of peptide collision induced dissociation (CID) spectra.

A quadrupole consists of four parallel rods, and opposite electrodes are connected to them. The ions pass lengthwise between the rods. Two opposing rods have both AC and DC voltages applied and the other two rods have the same potential carrying an opposite polarity. Oscillation of the applied voltages affects the trajectory of ions travelling through the flight path centred between the four rods. With the given DC

and AC voltages, only ions of a certain m/z ratio pass through the quadrupole filter. All other ions are eliminated by ejecting out of the path.

In MS-MS systems there are three quadrupoles. The first MS (MS-1) employs a quadrupole mass filter and allows only the selected analyte ions of interest to pass through. These selected ions are then taken to the collision cell where this ion is fragmented under an inert gas flow (argon) by application of a mild RF voltage, and the released so-called daughter ions are then swept into a second TOF analyser (MS-2) where they are separated and detected. The function of the first quadrupole (Q1) and the third quadrupole (Q3) is to filter ions of specific m/z ratio by scanning the potential applied to the quadrupole rods. Q2 is the collision cell and transmits ions without selection. Q3 records the m/z ratios of the fragment ions that originate from the fragmentation of the parent ion selected by Q1 (Figure 1.3). The ability to conduct tandem MS after electrospray ionisation has revolutionized proteomics since this MS-MS method can be used to obtain partial amino acid sequence information [99].

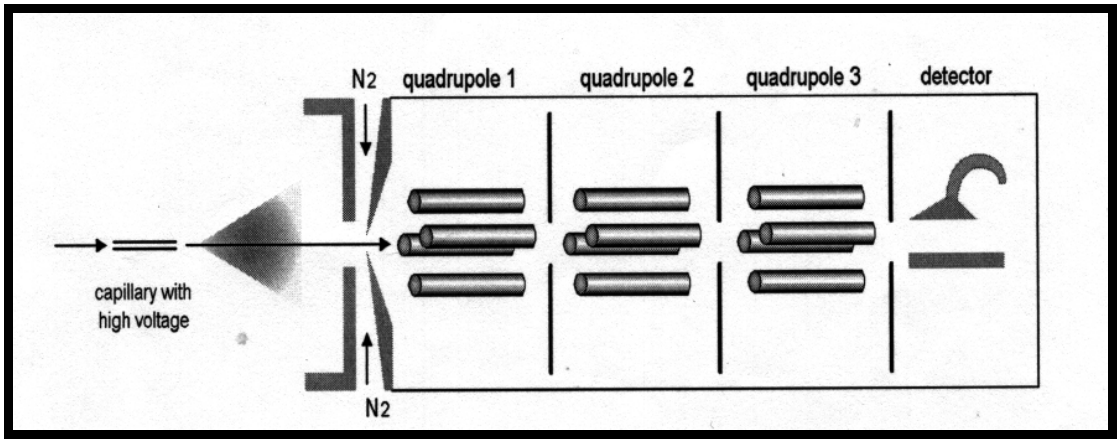

Figure.1.3 MS/MS Quadrupole system

1.4.2.4.3 TIME-OF-FLIGHT (TOF) MASS ANALYSER

Time-Of-Flight (TOF)

The first Time-Of-Flight (TOF) construction has been published in 1946 [100]. The principle of TOF is relatively simple. In TOF analysers, the mass to charge ratio of the ions are determined by measuring their flight time through a long tube. The ions of different m/z ratio are dispersed in time while they fly along a field free drift path of a known length. The ions start flying within a sufficiently short time interval; the smaller molecules arrive earlier at the detector than the heavier ones. At the end of

the flight the detector produces a signal for each ion species. Flying times are between a few microseconds to several hundred microseconds.

There are two kinds of TOF analysers, which are “Linear TOF” and the “Reflector TOF”. With the usage of reflected mode, the peak broadening is reduced. The ions are decelerated in the reflectron and turn around at different locations in the reflected electric potential gradient, thus ions of higher kinetic energy spend a longer time in the ion mirror. When the voltages and the position of the mirror are arranged accordingly, the arrival time spread can be corrected at the plane of the detector, and increased mass resolution can be obtained. In figure 4.1, the schematic representation of reflected mode of TOF is shown. The performance of TOF analysers has tremendously increased throughout the years [101, 102], which adapted the usage of TOF analysers with other ionisation methods. These hybrid instruments are now competitors with the well-established magnetic sector field instruments in many applications [103]. Advantages of using TOF analysers are that, they give rise to higher sensitivities, are cheaper, provide a complete spectrum within several tens of microseconds and most importantly in theory the m/z ratio is unlimited [104, 105].

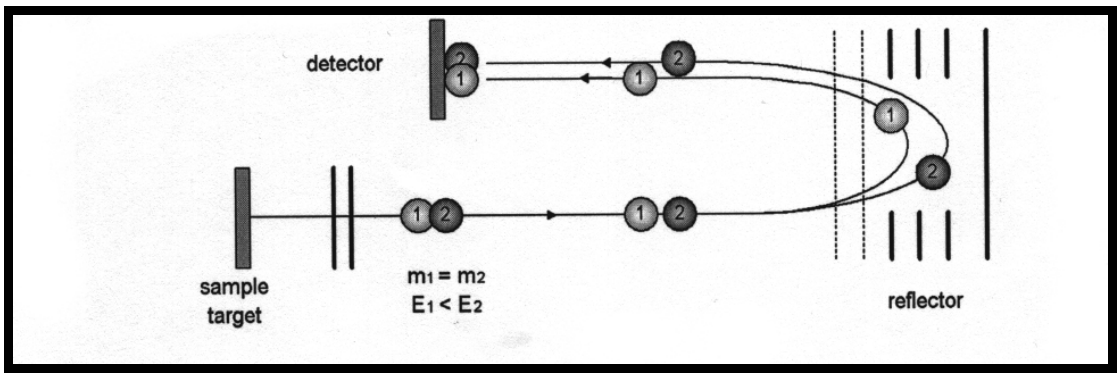

Figure.1.4 Reflected mode TOF analyser

1.4.2.4.4 NANO-LC-ESI-Q TOF: A HYBRID INSTRUMENT

The first hybrid Q-TOF instrument was developed to analyse peptides [106, 107], which is now used for many other applications like pharmaceutical and food analyses.

The Q-TOF is a hybrid quadrupole time of flight mass analyser with MS/MS capability. The quadrupole is operated as an ion guide in MS mode and as mass selection service in the MS/MS modes. A collision cell is located in between the quadrupole and the TOF analyser and induces fragmentation in MS/MS. The final

detector is a microchannel plate with high sensitivity. A nanoflow electrospray interface is used for small amounts of digested peptides derived from 2-DE gel spots. NanoLC Q-TOF is represented below (Figure 1.5).

The Q-TOF has very high sensitivity, very high resolution and precise mass accuracy. The high resolving power enables improved mass accuracy for peptides, charge state identification of multiple charged ions, and greater differentiation of isobaric species. This high mass measurement accuracy allows precise mass measurements of small molecules and *de novo* peptide sequencing from MS/MS spectra. With a nanospray source, the Q-TOF can analyse small sample amounts and may identify proteins through semi- and complete *de novo* sequencing.

Peptides derived from tryptic digest of proteins are particularly suitable for sequencing by the Q-TOF, largely because of the size of the peptides that vary between 800 to 2500 Da.

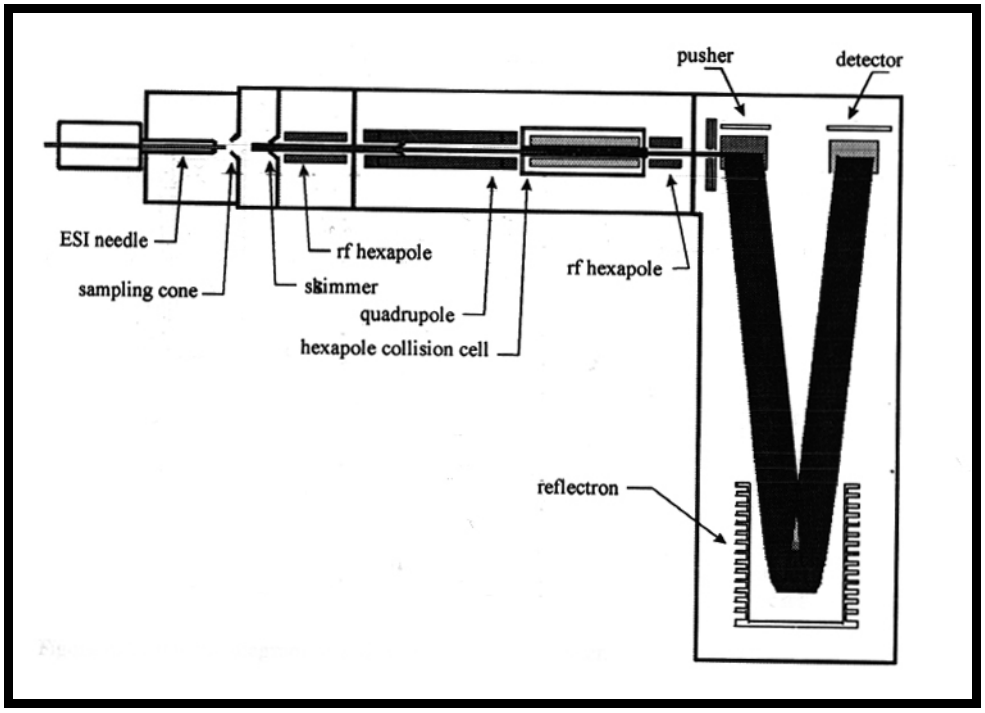

Figure 1.5 NANO-LC-ESI-Q TOF

1.4.2.4.5 THE DETECTOR

Electron multipliers and microchannel plates are used for detecting the ions in most of the mass spectrometers. Transmitted ions are deflected into the collector of the

detector. An ion striking the collector causes an electron cascade and the resulting signal after amplification is sent to the data acquisition of a computer.

1.4.2.5 DATA ANALYSIS

In most of the newly developed mass spectrometers for protein identification data analysis is performed by software for searching the masses of peptides within suitable databases.

For protein identification public genomic and protein databases are screened for an exact match. In the case of a novel protein, which is not contained within these databases, the *de novo* peptide sequence can be used for a BLAST search (Basic Local Alignment Search Tool). This may identify similar protein from another species, whose sequence can be found in a database.

When the complete genome, e.g. of a bacterial species is known, protein identification is possible as all possible protein sequences are represented in the genomic/protein sequence database.

There are two types of MS data used for protein identification by correlation with sequence databases. The first one is the determination of the accurate mass of peptides obtained by a specific cleavage of the isolated protein. MALDI-MS-TOF and ESI-MS produce data for this kind of data analysis. The collection of the masses of peptides generated from a digestion of a pure protein is correlated with sequence databases. This technique is known as the peptide mass fingerprinting (PMF). It is the identification of proteins already contained in a sequence database using an algorithm to match a set of peptide masses generated from a protein of interest. When the cleavage of the protein is mentioned in the search, theoretically peptide masses calculated from each sequence entry in the database are correlated to the experimental masses. A ranking is calculated to provide a measure of the fit between the theoretical and experimental masses.

The second type is the non-interpreted or partially interpreted MS/MS spectra data analysis. As the peptides are fragmented further by Q-TOF-MS/MS detailed sequence specific information can be obtained. Since each peptide will generate a sequence specific MS/MS spectrum, only a few spectra (minimally one spectrum) might be suitable to identify a protein. The help of programs that work interactive with the spectrometer helps tremendously to identify a protein since they are faster and conclusive. The information of the MS/MS spectrum gives very detailed information since it includes the mass of the intact peptide; sequence specific

fragment ions derived from the N-terminus (b series) and also the fragment ions from the C-terminus (y series).